Project Report
ATC-440

Correlated Encounter Model for Cooperative Aircraft in the National Airspace System Version 2.0

N.K. Underhill
E.P. Harkleroad
R.E. Guendel
A.J. Weinert
D.E. Maki
M.W.M. Edwards

8 May 2018

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

This material is based upon work supported under
Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001.

This material is based upon work supported under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.

Massachusetts Institute of Technology
Lincoln Laboratory

Correlated Encounter Model for Cooperative Aircraft in the National Airspace System
Version 2.0

N.K. Underhill
E.P. Harkleroad
R.E. Guendel
D.E. Maki
M.W.M. Edwards
Group 42

A.J. Weinert
Group 44

Project Report ATC-440

8 May 2018

Lexington Massachusetts

This page intentionally left blank.

ABSTRACT

This report documents an update to the 2008 Correlated Encounter Model for Cooperative Aircraft (CEM) subsequently referred to as the Extended Correlated Encounter Model (ECEM). This model generates statistically representative encounters between two aircraft in the U.S. National Airspace System (NAS) in which both aircraft are cooperative (carrying a transponder) and at least one is receiving Air Traffic Control (ATC) services. ATC involvement means maneuvers of one aircraft throughout the encounter are not wholly independent of the other–they are correlated. This model extends the encounter duration from the original 50 seconds to 120 seconds. Initial research on extending the CEM produced unrealistic behaviors and therefore several additional changes were implemented to prevent or reduce these behaviors and to achieve other identified potential improvments. The resulting ECEM produces high-quality, 120 second encounters that provide the increased time and distance required for testing Detect and Avoid (DAA) systems. The ECEM is one of many encounter models developed by MIT Lincoln Laboratory (MIT LL); other MIT LL designed encounter models are described in separate reports.

This encounter model was developed to support ongoing research and testing of collision avoidance and DAA systems, and specifically is designed to enabling testing of DAA systems supporting "Remain Well Clear" capabilities. The model was constructed using four months of radar data collected across the United States of America (USA) from 180 ATC radars. These data were used to build a Bayesian network with distributions and relationships identified from the encounters observed within those data. A dynamic Bayesian network is utilized to identify the statistical structure of maneuvers and subsequently produce maneuver behavior using those relations.

The ECEM is similar in structure to the CEM. It allows horizontal and vertical maneuvers, however, because of the extended duration it additionally allows acceleration variations. The ECEM also explicitly models airspeeds, vertical rates, turn rates, altitude layer, and airspace class enabling specific encounter subsets to be generated by limiting or specifying variable values in the encounter model structure.

This report documents data collection, data cleaning, the model framework, targeted model improvements, model validation, and encounter generation.

This encounter model and the statistical data to generate it are available from MIT LL in digital form. Descriptions, instructions, and examples for these files are provided in Appendices A, B, and C. Software to generate encounters using the encounter model is also available.

An additional byproduct of this effort is a database of recent, NAS wide encounters. These encounters can be used to enhance the existing encounter model, and/or refresh or build alternate encounter models with current data. These datasets can only be provided in electronic form due to their size.

This page intentionally left blank.

ACKNOWLEDGMENTS

The authors would like to thank Hill Air Force Base for access to radar data that was critical in the development of this model.

This page intentionally left blank.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

This page intentionally left blank.

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF TABLES

LIST OF TABLES
(Continued)

1. INTRODUCTION

Detect and Avoid (DAA) systems are designed to enable safe operation of Unmanned aircraft system (UAS) in the U.S. National Airspace System (NAS). The Radio Technical Commission for Aeronautics (RTCA) has been developing definitions and requirements for UAS DAA systems towards certification and mandate as Traffic Alert and Collision Avoidance System (TCAS) systems were for large transport aircraft [1]. DAA would actively prevent loss of separation (aircraft coming within 1,000 feet vertically and five Nautical Miles (NM) horizontally of each other) by proactively identifying, tracking, and providing guidance to avoid other aircraft, enabling the UAS to remain "well clear" of other air traffic.

As with TCAS, testing DAA systems before fielding is necessary to validate system robustness and confirm the system's safety benefit. These tests must be rigorous, ideally encompassing the full range of scenarios that might occur and would be addressed by a DAA system. These scenarios, referred to as "encounters," are situations where a loss of standard separation between two aircraft is likely or has occurred. Since testing should confirm proper functioning of basic system components prior to attempting tests involving many components, DAA system logic should be thoroughly tested through simulation of encounter scenarios to identify benefits, capabilities, and weaknesses prior to more complex and expensive evaluations such as flight testing.

Merely using observed flight track data is insufficient to achieve the diversity of situations a DAA system will need to address and therefore needs to be tested on. Situations where a loss of separation occurs are rare, resulting in the need for excessive amounts of data to be collected and processed to achieve both encounter quantity and diversity. Additionally, observed encounters will not span the full spectrum of possible scenarios.

One solution to achieve sufficient encounter quantity and diversity is to model the flight track data and relationship information in observed encounters and leverage the data and structural information to generate synthetic encounters. A model that sources its distributions and relationships from observed encounters enables generation of synthetic encounters that are realistic. However, generation will produce slight variations within the range of observed behavior enabling the creation of new encounters that have not been observed. This variation, bounded by sensible limits, enables generated synthetic encounter sets to expand to the full space of possible encounters that are within the range of observed encounter behavior. This approach was used to rigorously test TCAS, using millions of synthetic encounters to determine the capability of the system to reduce the risk of Near mid-air collision (NMAC) [2, 3]. These tests contributed to the eventual certification and U.S. mandate for TCAS equipage on larger transport aircraft. Testing TCAS with synthetic encounters was repeated by European Organization for the Safety of Air Navigation (EUROCONTROL) and the International Civil Aviation Organization (ICAO) for European and worldwide TCAS mandates [4, 5].

Rigorous testing of a system requires that the encounters used to test it are appropriate and accurate representations of what the system will be expected to address. Explicitly deconstructing observed encounters based on certain characteristics and subsequently building a model with those features enabled targeted generation of certain types of encounters where the controlled variables are set to desired values or limited to specific ranges. This approach has been used to develop many different encounter models from radar data, each able to create a diverse set of encounter scenarios. Models developed previously include those for cooperative traffic, noncooperative traffic, helicopters, oceanic encounters, and littoral encounters [6–10].

The 2008 Correlated Encounter Model for Cooperative Aircraft (CEM) was designed to test collision avoidance systems that detect conflicts 30 seconds seconds to one minute in the future. Remaining "well clear" requires DAA systems to detect a potential conflict up to 110 seconds in the future. This exceeds the duration of the encounters generated by the CEM, making it inadequate for testing DAA systems. To test DAA systems the Extended Correlated Encounter Model (ECEM), an extension of the CEM was developed. This report describes the ECEM, which generates encounters with trajectories beginning 110 seconds prior to Closest Point of Approach (CPA), enabling testing of the largest DAA alerting conditions that have been defined by the RTCA. Similar to the CEM, the ECEM uses a Bayesian network to capture encounter feature correlations and uses a Markov model to generate sensible and smooth aircraft maneuvers. The ECEM also contains some structural improvements over the CEM to better approximate aircraft trajectories over the longer time horizon.

2. APPROACH

The encounters generated by this model involve aircraft whose trajectories would likely trigger a collision avoidance or DAA system. We assume the standard separation distance has not been maintained despite existing safety mechanisms both structural (e.g., airspace structure) and operational (e.g., Air Traffic Control (ATC)), therefore creating scenarios where a DAA or collision avoidance system would be necessary to identify and resolve the situation. This model does not seek to address, nor is it appropriate to be used for other aircraft separation aspects such as ATC communication and coordination or airspace structure.

Encounters involving two aircraft are by far the most prevalent and are the only types used to build the model and the only type generated by the model. Aircraft trajectories used to build this model are also limited to aircraft using a transponder. There are two categories of aircraft represented based on transponder usage:

- Cooperative: The aircraft has a transponder and is transmitting a beacon code (Mode 3/A code).
- Noncooperative: The aircraft is not using a transponder and therefore not transmitting a beacon code.

Only cooperative aircraft were used to form this model.

Based on the beacon code an aircraft is transmitting, we can determine whether or not an aircraft was in communication with and therefore advised by ATC which would impact the anticipated behavior of the flight:

- Discrete Code: The aircraft is receiving ATC services. This includes aircraft flying under Instrument Flight Rules (IFR) and aircraft flying under Visual flight rules (VFR) but receiving flight following or other ATC services. These flights are identified by the discrete beacon code assigned by ATC.
- 1200 Code: The aircraft in the United States of America (USA) is flying under VFR without flight following and is not receiving an ATC service.

An encounter where at least one aircraft is a discrete code aircraft is considered to be a correlated encounter. An aircraft in communication ATC is assumed to exhibit maneuvers that are correlated due to ATC identifying a potential loss of separation and advising the aircraft in advance to prevent it. This encounter model focuses on correlated encounters; therefore only observed encounters with at least a single discrete code aircraft are used to build the model and all encounters generated by the model have at least a single discrete code aircraft as one of the pair.

Finally, encounters in this model represent only conventional aircraft, which are defined as aircraft likely carrying transponder equipment. Conventional aircraft include fixed-wing powered aircraft, whereas unconventional aircraft include balloons, blimps, gliders, and helicopters.

The ECEM is the only model described in this report; however, MIT Lincoln Laboratory (MIT LL) has developed a number of other encounter models to simulate different aircraft and encounter types including:

- Correlated Encounter Model [8]
- Uncorrelated Encounter Model [6, 7]
- Encounter Models for Unconventional Aircraft [9]
- Due Regard Encounter Model [10]

2.1 ENCOUNTER MODEL

An encounter is a situation where aircraft have failed to or are in danger of failing to "Remain Well Clear" of each other. Encounters are identified within radar data by evaluating aircraft pairs using an extended threat declaration logic. This results in trajectories of significantly separated aircraft qualifying as encounters. However, it is ideal to initialize encounters with sufficeient distance between the aircraft to enable the aircraft's DAA or collision avoidance logic to proceed through the full sequence of responding to an encounter: identifying, tracking, and resolving the threat. Although few observed encounters result in an NMAC, (two aircraft coming within 100 feet vertically and 500 feet horizontall of each other), generating and evaluating non-NMAC trajectories is necessary to check that a DAA system will not induce a loss of separation or worse, a collision.

An encounter is generated by the model in three steps: initialization, trajectory generation, and alignment. In initialization the two aircraft starting states and the relative position of the aircraft at Time of Closest Approach (TCA) are sampled from the initialization Bayesian Network. During trajectory generation, the initialized aircraft states are used as the starting state in a dynamic Bayesian network that iteratively samples subsequent trajectory variables for each aircraft. In alignment, the two trajectories are aligned spatially at TCA using the relative positioning information sampled during the initialization step.

2.2 BAYESIAN NETWORK STRUCTURE

The initialization step and the trajectory generation step both sample variables that are correlated to each other. For instance, it is unlikely to see an aircraft performing both a severe turn and a strong climb simultaneously. Similarly, only airspace classes that exist within the altitude range chosen must be possible to sample from. Therefore sampling variables independently could create encounter states not only physically unrealistic for the aircraft but also structurally impossible in regards to airspace. To capture these key dependencies, variables in the initialization and the trajectory generation steps are determined using Bayesian networks.

A Bayesian network is composed of nodes (the system variables, represented by rectangles) and directed arcs (indicating dependency relationships between variables, represented by arrows). Figure 1 is a diagram of the Bayesian network used for initialization and Figure 2 is a diagram of

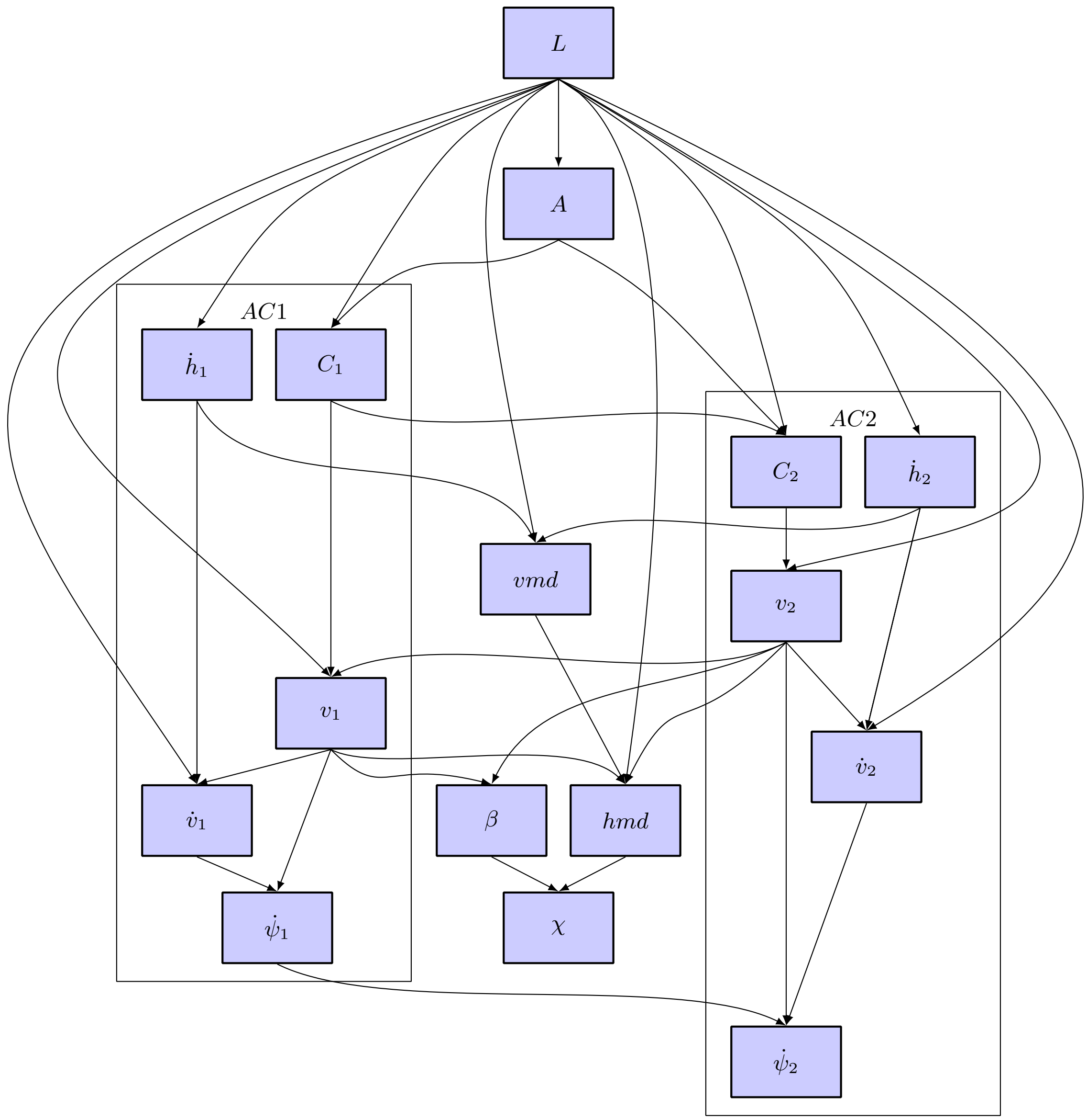

Figure 1. The Bayesian Network for the initialization step. Rectangles represent system variables, arrows indicate dependencies between variables.

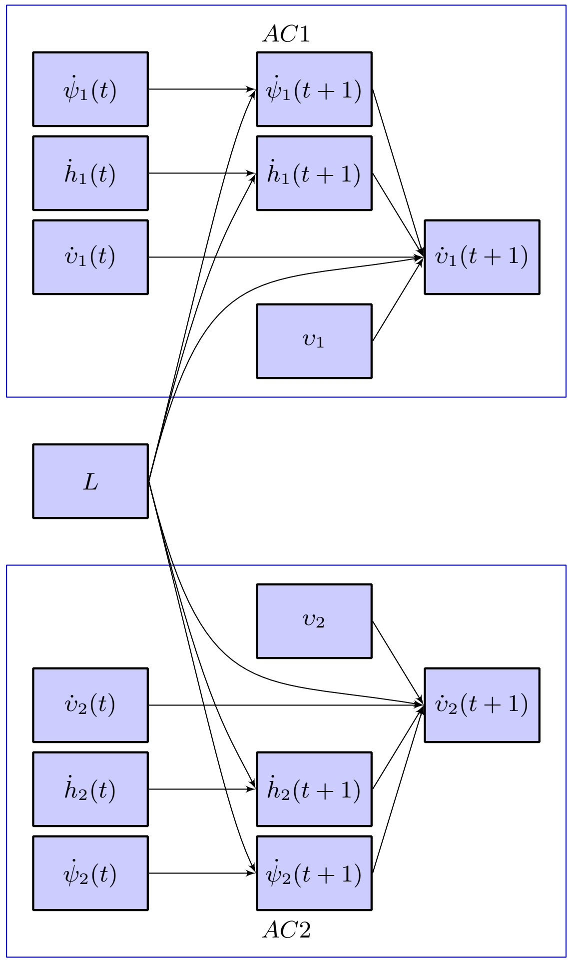

Figure 2. The Bayesian Network for the trajectory generation step. Variables with (t) represent variables at the current moment, variables with (t+1) represent variables at the next time step.

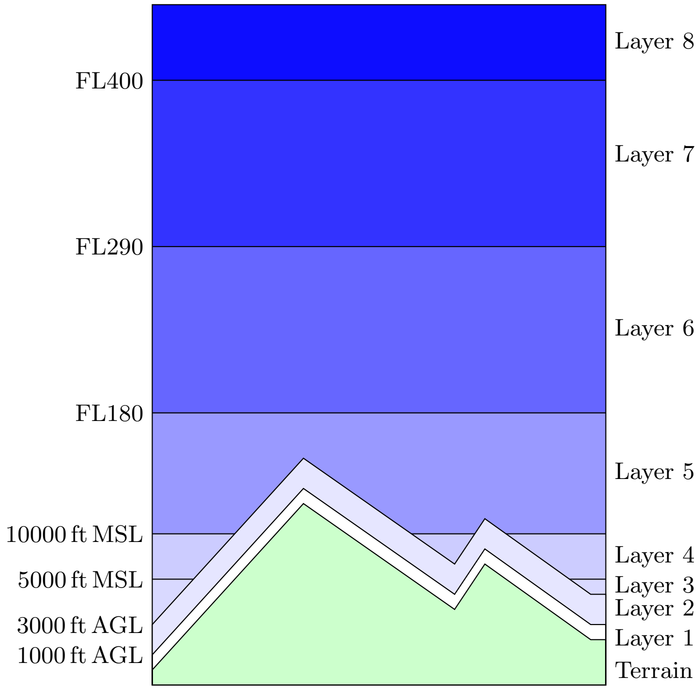

Figure 3. Altitude Layer quantization by MSL and AGL for the ECEM.

the dynamic Bayesian network used for trajectory generation. In a Bayesian network each node represents a single variable, each variable can take a single value, and each value has an associated probability of being selected. The directed arcs indicate which nodes have their probabilities influenced by the values of other variables. For example there is an arc from Altitude Layer L to Airspace Class A indicating that for each value of L there is a different probability for each possible value of A. For each node to have their value selected, all of the node's parent variables' values must already be determined. The variable vmd is a child of L, $\dot{h}_1$, and $\dot{h}_2$, and is therefore dependent on each; when sampling from the network all three of those variables must have their values determined before vmd can be set.

It would however, be computationally expensive and inflexible for the Bayesian Network to have an associated probability for every possible variable value. Limiting model complexity therefore demands a discrete set of values for each node; however preservation of variation (and model accuracy) requires representation of the full ranges of possible values. The solution is to quantize a variable's possible values in a logical fashion that enables thresholds and relationships to be identified and upheld. Retrospectively, some variables have already been quantized; for example Altitude Layer is a variable with a large range of values which have have already reduced been to a few specific ranges. Figure 3 shows the quantization of Altitude Layers.

The remaining continuous variables are quantized by binning their ranges into groups by cutting the range at multiple points. Each bin is assigned an associated probability of being selected through the Bayesian Network. Controlling model complexity and preventing overfitting can be easily managed with this method since the number and characteristics of bins are controlled the model parameters. In encounter generation, the model bin is sampled for a variable; the discrete value the variable takes within the bin is selected later.

The initialization step samples the initialization network once to determine all variable values, however the trajectory generation step requires a more complex process. This encounter model simulates nominal flight (without avoidance maneuvering) using a Markov process represented by a dynamic Bayesian network. Markov processes are stochastic methods where the probability distributions over future states are conditionally independent of past states, therefore only the present state is needed to predict the next state. A Markov model can be represented by an exhaustive state-transition matrix with specified probabilities for transitions between all state pairs. However, the number of parameters (probabilities) needed to define the matrix grows super-exponentially as the number of variables defining the model increase. Additionally, the more parameters used in the model the more data is needed to properly estimate their values. With dynamic Bayesian networks however, the conditional independence between some variables can be leveraged, greatly reducing the number of necessary parameters. The structure of the dynamic Bayesian network can be learned by maximizing the posterior probability of the network structure given the data.

2.2.1 Step 1. Initialization

In the initialization step the Bayesian network is used to sample the values that determine the overall characteristics of the encounter including: the airspace characteristics, the aircrafts' starting states, and the aircrafts' relative position at TCA. This initialization Bayesian network, illustrated in Figure 1, has the same nodes and directed arcs as the CEM and is evaluated once per encounter.

<u>Variables</u>

The initial Bayesian network contains four categories of variables; each variable category and their variables are described in detail below.

Encounter Location Variables: These variables describe location characteristics that impact the encounter characteristics. These variables are measured at TCA in the observed data, but are assumed to be true for the full duration of the generated trajectories.

> Altitude Layer (L): Altitudes are stratified based on prior research for the EUROCONTROL models and the CEM that looked at airspace structure (including airspace class). Layers are ranges of altitudes defined by either AGL or MSL elevations.
>
> Airspace Class (A): Airspace class is determined by altitude and proximity to airports. However, since altitude layer is already determined this variable only reports if the encounter occurs in airspace classes associated with airport proximity: B, C, D, and O for other airspace.
>
> Class A airspace can be identified by combining Altitude Layer and Airspace Class. Within Airspace Class O all altitudes above 18,000 feet MSL and below Flight Level (FL) 600 is Class A airspace. Class E and G airspace can be similarly identified.

Aircraft Type Variables: This variable describes whether ATC services are being provided to the aircraft. Values are determined through evaluation of the full observed trajectory and are consistent throughout the full generated encounter.

Aircraft Category (C_1, C_2): Aircraft type denotes whether the aircraft is a 1200 code aircraft, or a discrete code aircraft.

Trajectory defining Variables: These variables describe motion of the aircraft at the first point in the encounter.

Airspeed (v_1, v_2): The initial airspeed of Aircraft 1 (AC1) and Aircraft 2 (AC2).

Vertical Rate ($\dot{h}_1$, $\dot{h}_2$): The initial ascent or descent rate of each aircraft.

Turn Rate ($\dot{\psi}_1$, $\dot{\psi}_2$): The initial turn rate of each aircraft.

Acceleration ($\dot{v}_1$, $\dot{v}_2$): The initial acceleration or deceleration rate of each aircraft.

Relative Aircraft Position Variables: These variables describe how AC1 and AC2 are aligned relative to each other at TCA. These are measured at TCA during the observed encounters and only apply at TCA of the generated encounters. Figure 4 diagrams how these variables describe the relative position of the aircraft.

Approach Angle (β): This value is the heading of AC2 relative to AC1.

Bearing (χ): This is the bearing of AC2 relative to AC1.

Horizontal Miss Distance (hmd): The horizontal range between AC1 and AC2.

Vertical Miss Distance (vmd): The difference in altitude between AC1 and AC2. AC1 is always at the higher altitude, AC2 is at the lower altitude.

2.2.2 Step 2. Trajectory Generation

Initialization sets the starting conditions for both aircraft using a Bayesian network; the dynamic Bayesian network used during the trajectory generation step creates realistic trajectories for both aircraft. The transition network takes its starting state from the values generated in the initialization step and iteratively samples for each subsequent time step. Each set of trajectory values sampled determines the probabilities for the values of the next time step, resulting in realistic transitions between each set of trajectory values. Bayesian networks of this type which contain both current and future time states are known as dynamic Bayesian networks.

Bayesian networks can use any time interval between states; shorter intervals allow for more frequent variation but require relatively more computation, longer intervals limit variation frequency and are less computationally demanding. The ECEM uses a time step of one second to measure and generate high resolution maneuvers without imposing significant processing requirements.

Variables

There are two categories of variables in the transition network: constant and dynamic.

Constant Variables: These variables' values were determined in the initialization step; they are constant throughout the encounter duration but influence the generated variables. These include: Altitude Layer (L) and Airspeed (v_1, v_2).

Dynamic Variables: These variables are generated by the transition network and describe the movement of the aircraft over time. Each variable has a new value selected each second of the trajectory, with each value dependent on the values at the previous time step. This dependence prevents erratic, unrealistic changes in flight behavior and ensures generated trajectories will generally transition to the more significant turns, accelerations, climbs, and descents from their current state at a realistic and incremental rate. The trajectory generation or transition Bayesian network is successively used for each time step in the encounter after initialization which selects the values for the starting time. For the ECEM 120 second duration encounters the transition network is sampled 119 times with the values generated during each evaluation being used to sample variables at the next time step. The dynamic variables generated include: Vertical Rate ($\dot{h}_1$, $\dot{h}_2$), Turn Rate ($\dot{\psi}_1$, $\dot{\psi}_2$), and Acceleration ($\dot{v}_1$, $\dot{v}_2$).

The ECEM transition network is the CEM transition network with acceleration, and directed arcs for acceleration added. The transition network is diagramed in Figure 2.

2.2.3 Step 3: Alignment

In the alignment step the two trajectories created during the trajectory generation step are aligned at TCA using the relative position variables from the initialization step. For mathematical ease and consistency, AC1 is always facing due north at TCA and AC2 is aligned to AC1. Figure 4 diagrams how the two trajectories are aligned using the relative position variables generated during the initialization step.

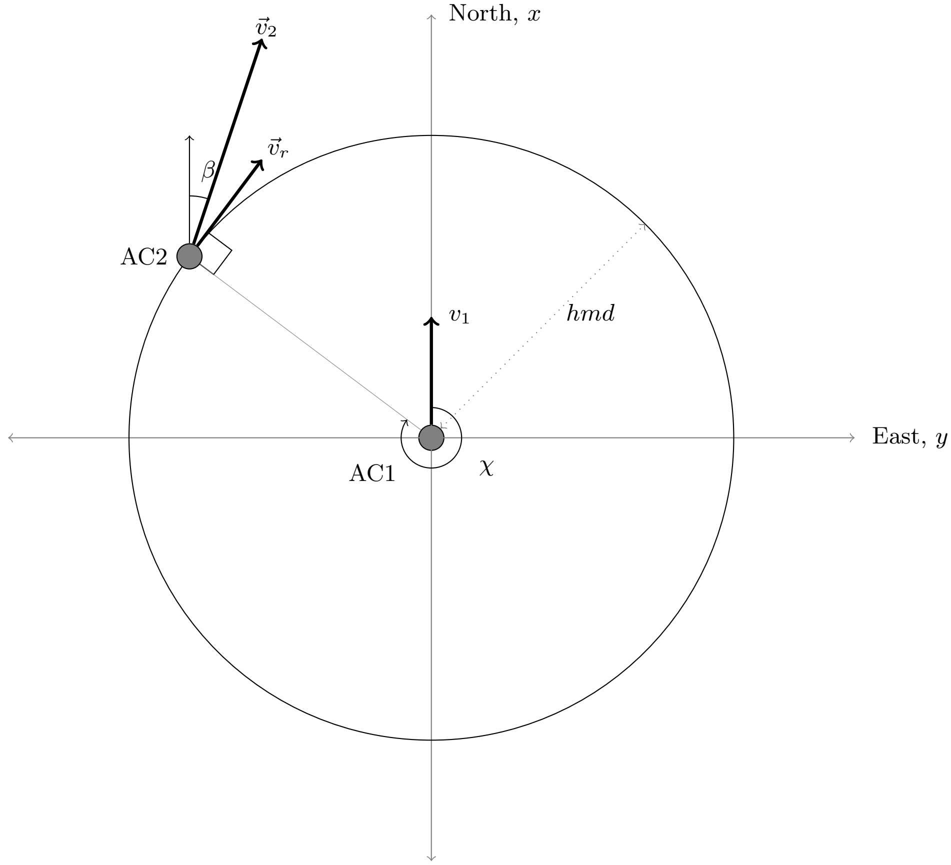

Figure 4. Approach angle (β) and bearing (χ) definition.

This page intentionally left blank.

3. ESTIMATION

3.1 DATA COLLECTION

Data for this encounter model was sourced from the 84th Radar Evaluation Squadron (RADES) at Hill Air Force Base, Utah. RADES data includes information from both Federal Aviation Administration (FAA) and Department of Defense (DOD) radar sites. Although RADES has data for 347 radars, due to processing constraints only data from 180 radars were used. A few examples of radars which provided information to RADES are listed in Table 1.

TABLE 1

Radars Contributing to RADES Data

Radar Name	Range	Scan Rate
ARSR-4	250 NM	1 / 12 seconds
ASR-8	60 NM	1 / 5 seconds
ASR-9	60 NM	1 / 5 seconds
ASR-11	60 NM	1 / 5 seconds

Air Route Surveillance Radar (ARSR)
Airport Surveillance Radar (ASR)
NM

A total of 299 days of RADES were processed. The ECEM targeted baseline was five million encounters contributing to the model. With this benchmark, 90 days of RADES data were used, including 5,760,000 encounters and over 512,000 hours of flight.

The full processed RADES dataset spanned the following dates:

- 2014: 30 Days (December)
- 2015: 190 Days (January–July)
- 2016: 79 Days (March–June)

Data from the following dates were used to build the ECEM:

- 2014: 30 Days (December)
- 2015: 60 Days (January–March)

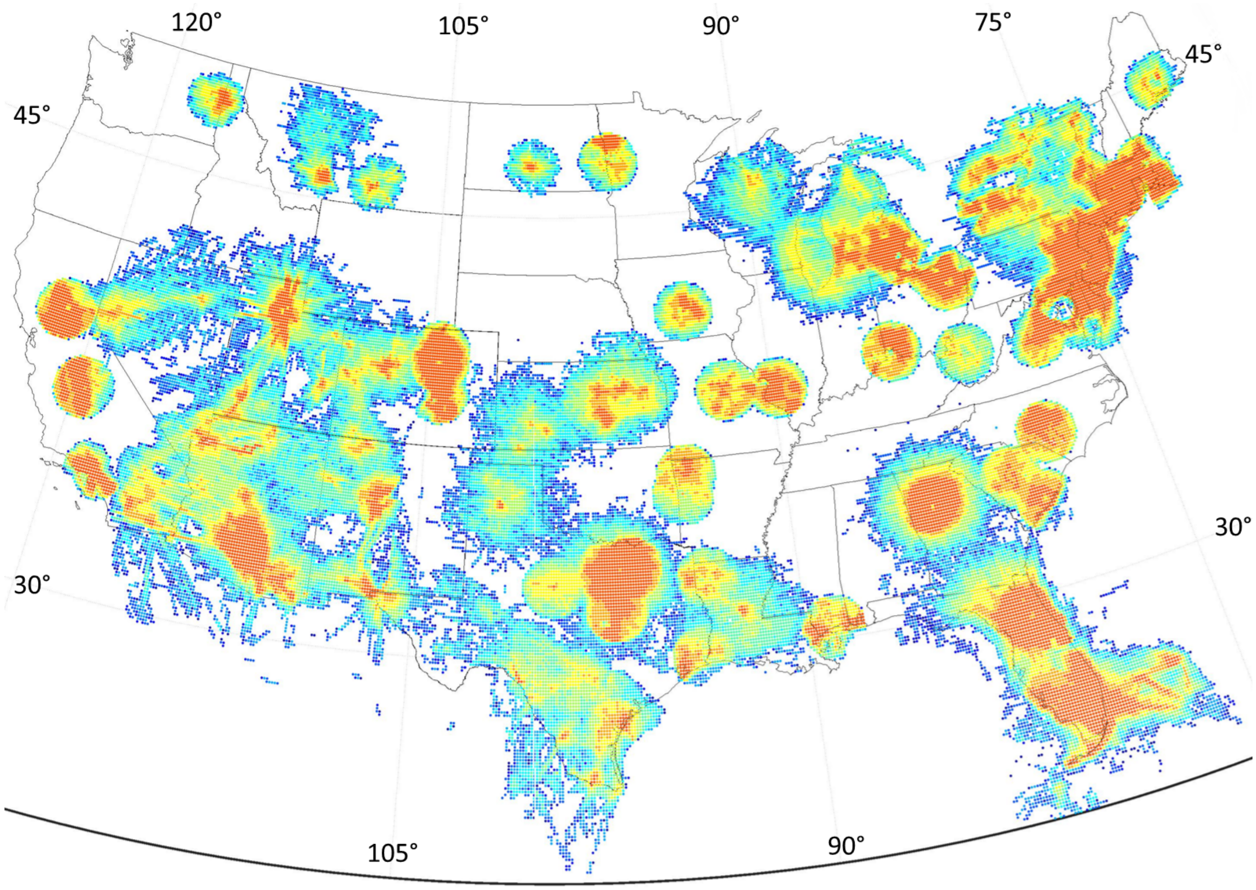

Figure 5. Density map of 1200-code aircraft.

3.2 ENCOUNTER IDENTIFICATION

Using a tracking algorithm developed at MIT LL tracks are formed from raw radar data. Filtering software subsequently evaluates each track's radar scans to check if it satisfies the criteria for an encounter with another track (Appendix E). Two tracks qualify as an encounter based on an alert-like test (Appendix D). If the encounter passes the alert-like test a single sensor's reports of the two tracks is saved into the encounter archive.

RADES data may have observations from multiple radars of the same track. However, minor differences (errors/biases/offsets) in radar azimuth and range measurements result in slightly different latitude and longitudes for the same flight. While using data from multiple radars can greatly increase a track's resolution, the data must be fused while maintaining the relative position and movement of both tracks. This fusion is computationally intensive, instead reports from a single radar is used for both tracks, maintaining relative position and movement while keeping processing streamlined and efficient.

3.3 DATA CLEANING AND FILTERING

After encounters have been identified, the associated data must be cleaned. This process takes the identified encounters and prepares them for ingestion into the model structure by removing noise and errors from the data prior to building the model. The following steps describe the process used to ensure only quality data is used in the model.

The data preparation step performs five distinct actions:

Data Quantity Requirements: Both encounter tracks must have suffecient data to accurately capture characteristics and maneuvers or the encounter is discarded.

Data Quality Disqualifications: Individual data points are evaluated for acceptability; extreme or erroneous values are discarded.

Encounter Requirement Disqualifications: An encounter must meet criteria for the encounters targeted by this modeling effort; encounter in military airspace, for example, are discarded.

Encounter Data Disqualifications: Track data outside the model relevant period is discarded.

Model Requirement Processing: Track data is smoothed and interpolated to achieve realistic and regular track representation.

3.3.1 Data Preparation Process

This section describes the cleaning and filering process for the encounter data. Steps are listed chronologically.

Initial Encounter Requirement: Each encounter must have two tracks from the same radar, the two tracks must have some overlap in time. Encounters that do not meet this requirement are discarded.

Initial Data Requirement: Track information from the same radar is collected for a period of ten minutes centered on TCA for both tracks. Each track must have at least ten data points. If either track fails this requirement the encounter is discarded.

Next, multiples of data points with the same time are removed from track data. Each tracks must then have a minimum of 32 data points over the ten minute period. (This results in an average resolution of one scan every ninteen seconds.) If either track has insuffiecient data the encounter is discarded.

Beacon Data Cleaning: When data points were identified as part of a trajectory, unrelated points that occur in close proximity to the track may be included. To remove this mistakenly included data, each track is evaluated using the following process:

1. All data points with a beacon code of 1200 with an altitude greater than zero are retained.
2. If the track includes discrete beacon codes, data associated with the most prevelent discrete beacon code are retained.
3. All other beacon codes and associated data are discarded.
4. If there are more 1200 codes than discrete codes the track is determined to be a 1200 code flight, otherwise it is discrete code flight.

Note that this allows for the inclusion of flights that either transition from 1200 code to discrete or vice versa.

Outlier Removal: Data are removed from the track information (along with their associated information) if they meet any of the outlier criteria described in Table 2.

TABLE 2

Data Outerlier Removal Criteria

Variable	Outlier Criteria	Reason
Altitude	= 0 MSL	Missing mode C altitude report.
Vertical Rate	$\geq$ 5,000 feet/minute $\leq$ -5,000 feet/minute	Extreme limits of a maneuvering aircraft.
True Airspeed	$\geq$ 800 knots	
Acceleration	$\geq$ 6 knots/second $\leq$ -6 knots/second	
Turn Rate	$\geq$ 10 degrees/second $\leq$ -10 degrees/second	

Figure 6 illustrates an instance of outlier data identified for removal.

Track Overlap Trimming: Only the period where both tracks have information is used. The two tracks' data are compared and only the overlapping time period of both the tracks is retained. If there is no time overlap between the two tracks, the encounter is discarded.

Initial Data Requirement (Second Pass): Two reductions of the data have been performed at this point. Prior to the more computationally intensive checks, another pass of the Initial Data Requirement check is performed. Both tracks must have 32 points of data within ten minutes centered on TCA or the encounter is discarded. This removes encounters which have fallen beneath the data requirement threshold due to reductions.

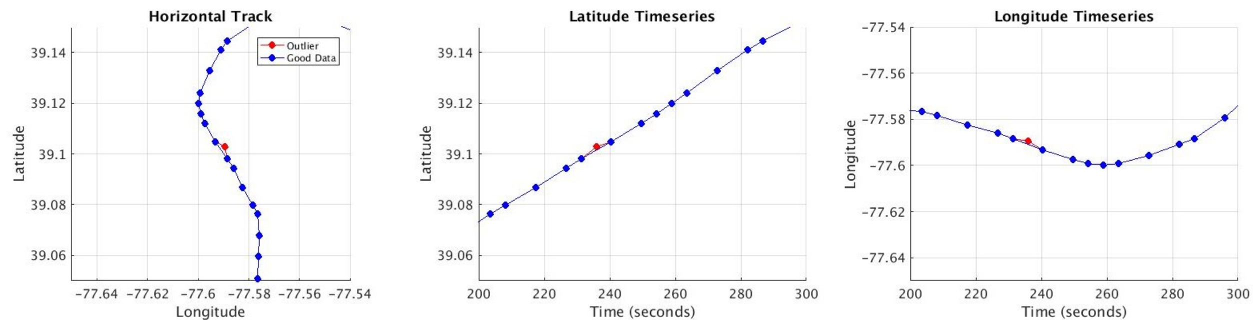

Figure 6. Horizontal track and timeseries graphs demonstrating outlier removal.

Horizontal and Vertical Smoothing: Smoothing is done using a locally weighted temporal smoother with a Gaussian kernel. Smoothing is performed independently on horizontal and vertical data; the width/standard deviation of the Gaussian kernel is five seconds for horizontal variables and fifteen seconds for vertical variables. Smoothing removes or reduces noise and ensures smooth transitions—preventing large jumps between very different values; instead flights will transition from straight flight to a slight turn to a stronger turn rather than changing directly to a strong turn from forward flight (which would not be realistic). Vertical and horizontal smoothing are illustrated in Figures 7 and 8 respectively.

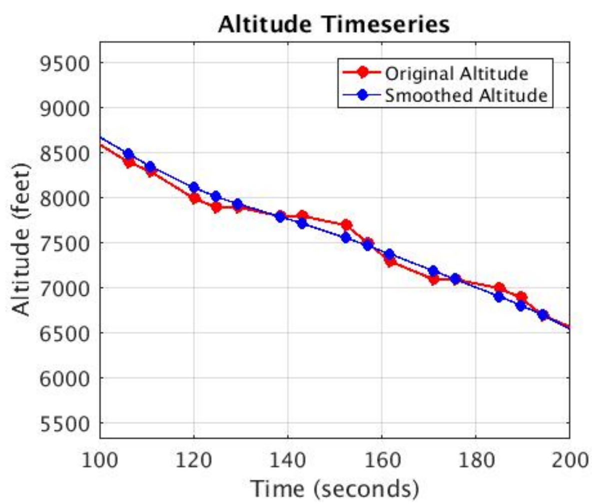

Figure 7. Smoothing of the track altitude.

Correlated Encounter Requirement: A correlated encounter requires at least one of the flights be in communication with ATC. A flight that is not recieving ATC services in the USA will squawk 1200. If both aircraft are squawking 1200, neither are in communication with ATC, the encounter is not correlated and is therefore discarded.

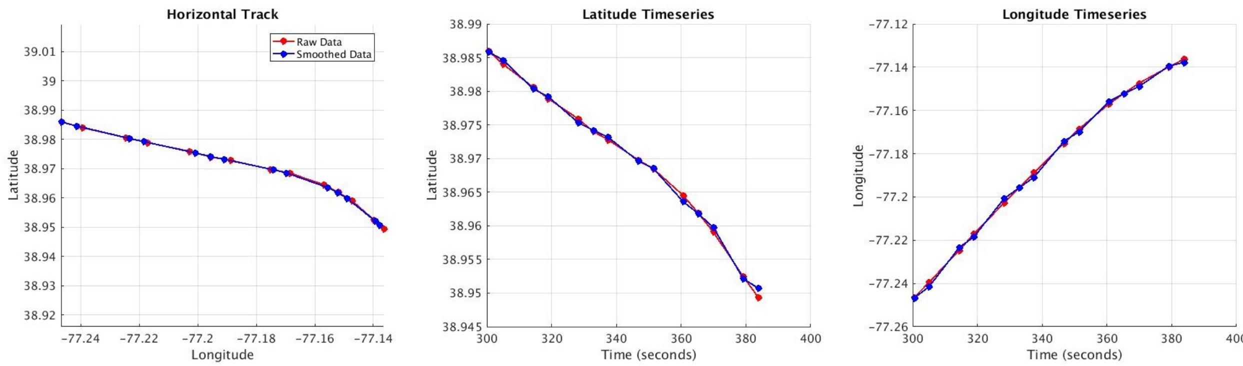

Figure 8. Smoothing of the horizontal track.

Stationary Track Encounter Disqualification: An encounter is unqualified if either track is stationary. A track is determined to be stationary if the cumulative distance traveled by the track is less than one NM. If either track is determined to be stationary the encounter is discarded.

Track Interpolation: The encounter model requires a one second track resolution, finer than is present in the radar data. To meet the increased resolution desired the track data is interpolated. The interpolation method used is the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) method. The PCHIP method does not overshoot and has less oscillation compared to a spline. PCHIP also preserves the shape of the data and upholds monotonicity when present [16]. Using PCHIP interpolation, all location and altitude values during the ten minute duration are replaced by the interpolated values.

Parallel Runway Encounter Disqualification: Parallel runways create conditions where aircraft appear to be in conflict due to close proximity and airspace or procedural structures–for example landing on parallel runways. Without precise comparison to existing and in–use procedures it is not discernable if the aircraft are in a potential conflict or not. Due to this ambiguity, encounters that occur in proximity to parallel runway airports are discarded.

Formation Flight Encounter Disqualification: Encounters where aircraft are intentionally in close proximity due to formation flying are also discarded. Formation flights can be flying within close enough proximity to count as a NMAC. Formation flights are identified by the similarity of their movements and maneuvers. Formation flights are not representative of a potential self-separation violation during standard flying practices and are discarded.

Trimming to Capture Period: Tracks are trimmed to the targeted period around TCA. For the ECEM the period for evaluation is 130 seconds prior to TCA and 30 seconds after TCA. Data outside the targeted range is discarded.

Final Data Requirement: Each encounter is maximally 160 seconds in duration and there must be at least 155 data points per track for the encounter to qualify as having sufficient data. This is a slight requirement increase from the CEM's requirement: 45 points for a 50 second duration encounter. This minimum data requirement ensures sufficient track information to accurately represent any maneuvers or movements that will be used to build the encounter model.

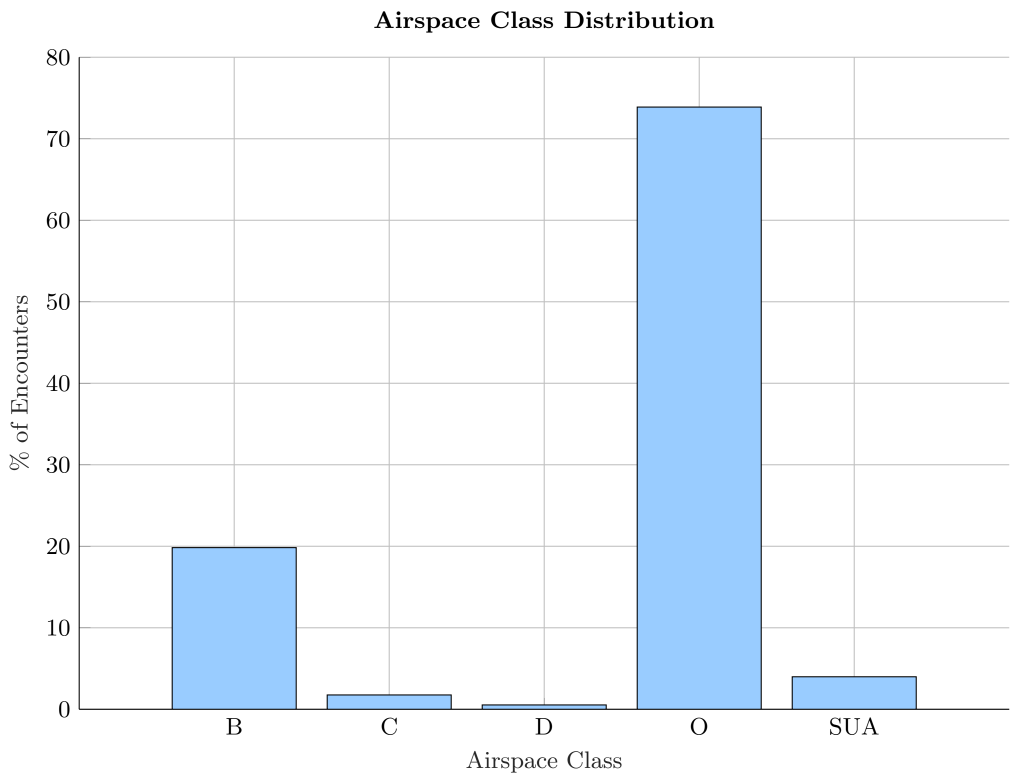

Figure 9. Histogram of airspace classess of observed encounters.

Mode C Report Requirement: This check requires a number of reports based on the scan rate of the sensor. The maximum time interval is assumed to be the radar's scan rate; the number of expected reports is derived from the scan rate, and the number of scans in the period around TCA (50 seconds before, ten seconds after) must be greater than 60% of the expected number of reports.

Special Use Airspace (SUA) Disqualification: Military flights are not appropriate encounters to build the model; military operations are not representative of standard airspace and may include tactical military training operations. Each encounter is evaluated to discern if the CPA occurs in SUA. Latitude, longitude, MSL, and AGL are used to determine which airspace class that the CPA occurs in. SUA data is sourced from the National Airspace System Resources database (NASR) provided by the FAA. SUA airspace is determined by the Digital Aeronautical Flight Information File provided by the Natioanl Geospatial Intelligence Agency (NGA). All SUA airspace is assumed to be active (although this is not always so) and all encounters with a CPA in SUA are discarded. Mode C reports are used to determine altitude (uncorrected for barometric variation so possibly inaccurate). Ground elevation is from the Digital Terrain Elevation Data (DTED) provided by the NGA. DTED level zero is used which has post spacing of 30 arc seconds (approximately 900 meters). Figure 9 shows the percentage of encounters disqualified due to CPA occuring in SUA. Note that airspace is only checked for at CPA.

This page intentionally left blank.

4. BUILDING THE MODEL

After the process described in the previous section, the remaining encounters are all of sufficient quality. To contribute to the statistical model, however, the characteristics and maneuvers must be extracted from the trajectories, categorized into quantized bins, and ingested into the model.

4.1 FEATURE MEASUREMENT

The variables that make up the model have already been defined in the Encounter Characteristics, Section 2.2. This section lays out the details of how each variable is measured from the processed observed encounters.

Airspace Class (A): The altitude, latitude, and longitude of AC1 at CPA are compared against the FAA NASR database to identify which airspace class the aircraft was in.

Altitude Layer (L): The altitude layer is determined by the AGL and MSL altitude of AC1 at CPA. MSL altitude produced by the processed radar scans, AGL altitude is estimated by subtracting ground elevation from MSL altitude. Ground elevation estimates are from DTED provided by the NGA. Figure 3 illustrates the Altitude Layers ranges by AGL and MSL.

Aircraft Category (C_1, C_2): The beacon codes for a flight are reduced to 1200 codes and the most prevalent discrete code. If there are more 1200 codes than discrete codes, the the flight is 1200 and an "1200 Code" otherwise it is a "Discrete Code" aircraft.

Approach Angle (β): The approach angle is the difference in headings between AC1 and AC2 at TCA. Aircraft headings are the arctangent of the ratio between the change in x and y of the aircraft at each step in time.

Bearing (χ): This is the bearing of AC2 relative to AC1 at TCA. Bearing is calculated as the arctangent of the ratio of the difference between the two x, y aircraft positions at TCA.

Horizontal Miss Distance (hmd): The horizontal miss distance is the distance between x and y of AC1 at TCA and x and y of AC2 at TCA.

Vertical Miss Distance (vmd): The vertical miss distance is the difference in altitude between AC1 and AC2 at TCA.

True Airspeed (v_1, v_2): Airspeed for time t is calculated by the following equation:

$$v(t) = \sqrt{(x(t+1) - x(t))^2 + (y(t+1) - y(t))^2 + (h(t+1) - h(t))^2}. \qquad (1)$$

Airspeed is smoothed using a locally weighted temporal smoother with a Gaussian kernel, the standard deviation is set to 20 seconds.

Acceleration ($\dot{v}_1$, $\dot{v}_2$): Acceleration is the difference in speed between the current second to the next. This is calculated using Airspeed that was calculated and smoothed immediately prior. Acceleration is smoothed with the same smoothing technique with the standard deviation set to fifteen seconds.

Turn rate ($\dot{\psi}_1$, $\dot{\psi}_2$): Turn rate at time t is the difference in heading at time t and the heading at time $t+1$. Turn rate is then smoothed with the same smoothing technique with the standard deviation set to ten seconds.

Vertical rate ($\dot{h}_1$, $\dot{h}_2$): The vertical rate at time t is the difference between the altitude at time t and the altitude at time $t+1$. Altitudes were smoothed during the data processing step so no additional smoothing is applied at this step.

4.2 DATA INGESTION

There are a few final steps before the encounters are ingested into the model.

Buffer Removal: At this point variable values are final and the buffer of additional dynamic trajectory variables is no longer needed. Variables within 20 seconds from the end of the trajectory are discarded, the 120 seconds of dynamic trajectory variables prior to that are kept, and all other trajectory data is discarded. Since the duration to this point has required 155 data points minimum all trajectories must be trimmed. Although the relative position variables were calculated precisely at TCA, this process induces a potential offset error of up to five seconds for the trajectory in alignment with TCA since it assumes missing values occurred the beginning of the trajectory.

Variable Value Quantization: All variable ranges must at this point have a quantization scheme. Airspace Class, and Aircraft Category have all already been quantized. The remaining variable quantization schemes are listed in Table 3. Two schemes are used: quantization along regular intervals, for example True Airspeed's range is cut into 50 knot intervals, and quantization due to meaningful value thresholds for example 500 feet for *hmd* from the NMAC definition. A point where the variable range is divided into separate bins is referred to as a "cut point."

Zero Bin: Continuous variables that span both positive and negative values such as turn rate, acceleration, and vertical rate have the range immediately around zero handled differently. Due to noise in radar scans and subsequent processing it is rare that aircraft are measured going perfectly straight, level, or maintaining speed. Instead, many small non-zero values are observed. This presents a problem; when the model is used for encounter generation the values within a bin

TABLE 3

Variable Range Cut Points

Variable	**Cut Points**	**Units**
Altitude Layer (L):	1000, 3000,	feet AGL
	5000, 10,000, 18,000, 29,000, 40,000	feet MSL
Approach Angle (β):	30, 60, ..., 330	degrees
Bearing (χ):	90, 270 (angles 270-360/0-90 are a bin)	degrees
True Airspeed (v_1, v_2):	100, 150, ..., 550	knots
Acceleration ($\dot{v}_1$, $\dot{v}_2$):	±4, ±3, ±2, ±1, ±0.25	knots/second
Vertical Rate ($\dot{h}_1$, $\dot{h}_2$):	±5000, ±4000, ±3000, ±2500, ±2000, ±1500, ±1000, ±400	feet/minute
Turn Rate ($\dot{\psi}_1$, $\dot{\psi}_2$):	±7, ±6, ±5, ±3.5, ±2, ±1, ±0.25	degrees/second
Horizontal Miss Distance (hmd):	500,	feet
	0.5, 1, 1.5, 2, 2.5	NM
Vertical Miss Distance (vmd):	100, 200, ..., 1000,	feet
	1200, 1400, ..., 5800	

will be uniformly distributed across the range of the bin resulting in few instance of consistent flight.

Using the quantization methods of the other variables would result in two options. If a bin was created with a small range centered on zero, it would ensure that generated values would be zero or very close but the number of observations due to the restriction would be few and as a result generating periods of steady flight would be similarly rare. Alternatively, a larger bin centered on zero would capture many more instances of noisy steady flight and the number of samples within this bin would be better representative of the frequency of constant conditions, however generation would create equally if not more noise for a bin that is targeted to represent unchanging conditions. Both of these solutions are unsatisfactory, so a new method for handling the zero ranges was developed.

A large bin is used to capture all observations that are likely noisy steady flight; however, when the model is used for generation the value produced by this bin will always be zero. This sidesteps the presence of noise, enables accurate frequency of steady conditions to be represented, and produces steady conditions in the encounter generation. These special conditions require that the bin containing zero for acceleration, turn rate, and climb rate be set conscientiously.

Variable Range Limits: Similar to the cut points, Variable Range Limits dictate the range that will be generated by the model.

Although cut points will divide the range of a variable the values accepted into the model are unbounded. Extreme values are counted towards the end bins of a variable for example a *vmd* of 7000 feet will be counted in the [5800,6000) bin. The range of these end bins must be set to determine how far those bins extend for instance the maximum value set for *vmd* will be 6000 feet. Table 4 lists the Variable Range Limits for the ECEM.

TABLE 4

Variable Range Limits

Variable	Minimum	Maximum	Units
Approach Angle (β):	0	360	degrees
True Airspeed (v_1, v_2):	50	600	knots
Acceleration ($\dot{v}_1$, $\dot{v}_2$):	-5	5	knots/second
Vertical Rate ($\dot{h}_1$, $\dot{h}_1$):	-6000	6000	feet/minute
Turn Rate ($\dot{\psi}_1$, $\dot{\psi}_2$):	-8	8	degrees/second
Horizontal Miss Distance (hmd):	0	3	NM
Vertical Miss Distance (vmd):	0	6000	feet

4.2.1 Network Representation

The counts for each bin are compiled into an array for each variable. Airspace Class A for example is shown in Table 5. A is a child variable of L, so the array in this instance has a column representing each Altitude Layer and a row representing each Airspace Class. The value in a row-column combination is the number of times that Airspace Class was observed within the specific Altitude Layer.

TABLE 5

Bin Counts for Airspace Class Given Altitude Layer

L	**B**	**C**	**D**	**O**
1	6238	3702	5066	5063
2	147282	58524	21803	232745
3	347529	32112	613	300295
4	497452	6872	3154	713507
5	144028	41	0	1309819
6	0	0	0	673961
7	0	0	0	963531
8	0	0	0	57328

This page intentionally left blank.

5. ENCOUNTER GENERATION

Once the data has been ingested into the ECEM structure, it can be used to generate new encounters representative of those used to build it. Encounter generation for the ECEM is identical to that of the CEM.

5.1 SAMPLING METHODOLOGY

While sampling from the Initialization and Trajectory Generation Networks have slight difference, most aspects of sampling are performed the same.

Sampling must be done in order. For a variable to be sampled all of its parent variables must have their bins already determined. Dependencies require that the Initialization Network be sampled before the Trajectory Generation Network. At the network level this means that the Initialization Network must start with sampling L and the Trajectory Generation Network always samples $\dot{v}_1$ and $\dot{v}_2$ last.

Next, a Dirichlet where the the prior distribution is uniform, with a bin count of one in all cases, is used as a base for determining bin probabilities. This ensures that there are no transitions with zero probability in the model, and also adds one to the counts within the bins during sampling. This will enable the generation of encounters not based in observed data if areas within the network have a zero in their bin counts.

Given the bins selected for the parent variables, the probabilities for the variable's bins are calculated. Table 5 shows the bin counts for which Airspace Class is likely to be chosen based on a given Altitude Layer. These values are added to the Dirichlet (adding one to all counts) and the probability percentages are achieved by dividing the counts for a given Altitude Layer by their combined sum. Equation (2a) shows how to calculate a particular bin probability percentage, Table 6 shows the probability percentages for all A given a particular L.

$$P(A = B|L = 1) = \frac{Count_{B|L=1} + 1}{(Count_{B|L=1} + 1) + (Count_{C|L=1} + 1) + (Count_{D|L=1} + 1) + (Count_{O|L=1} + 1)} \tag{2a}$$

$$P(A = B|L = 1) = \frac{6238 + 1}{(6238 + 1) + (3702 + 1) + (5066 + 1) + (5063 + 1)} \tag{2b}$$

$$P(A = B|L = 1) = \frac{6239}{20073} \tag{2c}$$

$$P(A = B|L = 1) = 0.3108 \tag{2d}$$

$$\tag{2e}$$

All variables except for A, C_1, C_2, and χ require subsampling to de-discretize within the bin selected (L must also be de-discretized however this occurs outside of the model). Subsampling within a bin is uniform across the bin range which is defined by the variable cut points (Table 3)

TABLE 6

Bin Probabilities for A Given L

L	**B**	**C**	**D**	**O**
1	0.3108	0.1845	0.2524	0.2523
2	0.3199	0.1271	0.0474	0.5056
3	0.5107	0.0472	0.0009	0.4413
4	0.4074	0.0056	0.0026	0.5844
5	0.0991	0.0000	0.0000	0.9009
6	0.0000	0.0000	0.0000	1.0000
7	0.0000	0.0000	0.0000	1.0000
8	0.0000	0.0000	0.0000	0.9999

and the variable range limits (Table 4). The ECEM additionally uses zero bins which rather than producing a uniform distribution over the bin spanning zero will return the value zero when that bin is selected. It is important to note that at this point variable range limits can be altered to change the range of the first and last bins. Also, zero bins can be turned off to produce a uniform distribution over the bin including zero.

Subsampling within the Trajectory Generation Network bins does not occur for every bin sampled; instead the subsampling occurs based on fixed probabilities. These probabilities are estimated from the radar data and assess how often track values change within a bin. Specifically bins are divided into three sub-bins and the probability of subsampling depends on how often a track within a bin will move between the sub-bins. EUROCONTROL's cooperative encounter model utilized a similar strategy. The probabilities for resampling Trajectory Generation Network variables are shown in Table 7.

TABLE 7

Resampling Rates for Trajectory Generation Dynamic Variables

Variable	Resampling Probability
Vertical Rate AC1 ($\dot{h}_1$):	0.0820
Vertical Rate AC1 ($\dot{h}_2$):	0.0655
Turn Rate AC1 ($\dot{\psi}_1$):	0.0898
Turn Rate AC2 ($\dot{\psi}_2$):	0.1035
Acceleration AC1 ($\dot{v}_1$):	0.0248
Acceleration AC2 ($\dot{v}_2$):	0.0227

This page intentionally left blank.

6. MODEL CHANGES AND UPDATES

An initial step of developing the ECEM was to run the CEM with the trajectory generation step extended to 120 seconds. This effort was followed by a thorough evaluation of the CEM's generated encounters in which a number of potential improvements were identified to be resolved in the ECEM. This section provides details of the implemented changes and the reasoning for their implementation including references to any Targeted Improvements (detailed in the next section) that they may have been implemented to resolve.

6.1 UPDATED SOURCE DATA

There were a few options when initially deciding on how the ECEM should be developed. Options included:

1. Rebuilding from the 2008 CEM dataset of cooperative correlated encounters.
2. Supplementing the 2008 CEM encounters with additional current information.
3. Using an entirely new dataset.

It was decided that the ECEM would be made using the final option, rebuild with only new data. The reasons behind this selection were:

> To achieve current encounter demographics for previously evaluated regions: the demographic of encounters may have changed since 2008. Recent data will capture encounters that are currently occurring. Encounters observed in older data may no longer be prevalent or relevant due to new procedures, or updated self-separation or collision avoidance technology.
>
> To incorporate encounters from new regions: recent RADES data included significantly more contributing radars than older RADES data enabling identification and inclusion of encounters in areas with insufficient coverage previously.
>
> To achieve a baseline number of encounters contributing to the model: the CEM started with about five million encounters. To develop the ECEM with a similar number of encounters, additional new encounters would have to be added to the existing set.
>
> To expand the set of encounters used to include those relevant for DAA, new identification criteria was developed. The new criteria is explained is explained in the next section.

Figure 10 shows the number of encounters identified per day processed by month and year. Figure 11 is a histogram showing the frequency and spread of encounters identified per day.

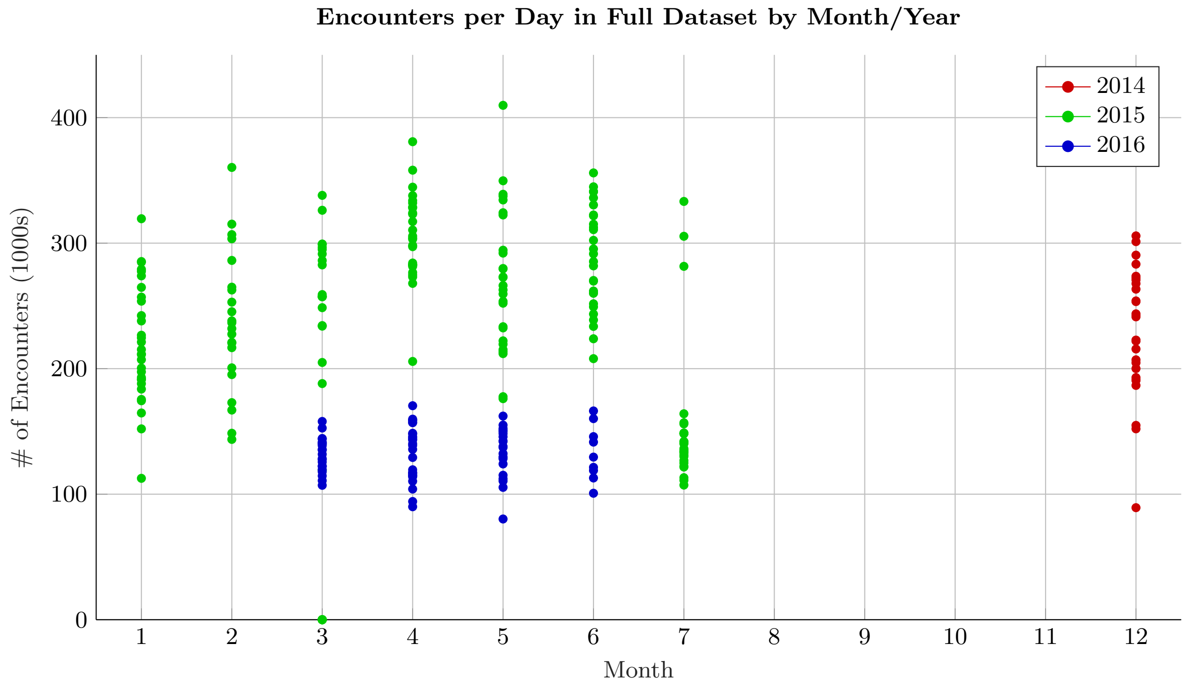

Figure 10. Encounters identified per day by date for all dates processed.

6.2 EXPANDED ENCOUNTER IDENTIFICATION CRITERIA

The ECEM will support collision avoidance and DAA system testing, and for this encounter identification using a TCAS-like test had to be updated from CEM assumptions to additionally include encounters necessary for testing DAA systems. The encounter qualification criteria were taken from the minimum alert conditions identified through analysis by the RTCA Special Committee 228: Minimum operational performance standards (MOPS) for UAS (SC-228) [1]. These criteria were then expanded with a 50% margin to ensure inclusion of all relevant encounters. Details of the filter are available in Appendix D. This resulted in an encounter set that included safely separated encounters as well to check that the threat resolution logic would not create or worsen a situation which had safe separation. This diversity of encounters is essential to thorough and rigorous testing of the system.

6.3 BUFFERED ENCOUNTER DURATION FOR PROCESSING

The data processing and features measurement steps assume a trajectory duration with a significant buffer before and after the actual period of interest. The ECEM is created with and generates 120 second trajectories, however processing and feature measurement occurs assuming a 160 second trajectory.

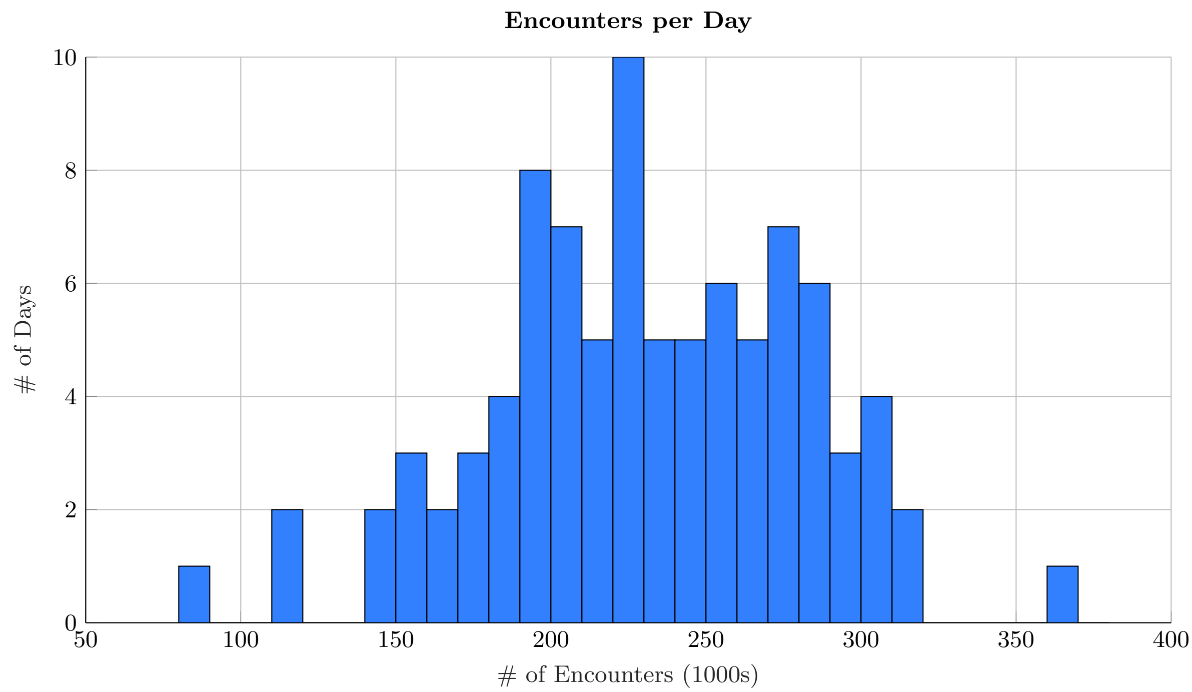

Figure 11. Histogram of encounters identified per day by date in the new RADES data.

Artifact Reduction: The reason for the increased encounter duration was to reduce or eliminate data artifacts. These artifacts are detailed in Targeted Improvements, Section 7. The duration of the buffers (20 seconds) and the location of the buffers are both directly to reduce artifacts.

Improved Encounter Quality: A secondary benefit was the increased accuracy and quality of encounters. Processing buffered encounters improves the accuracy of the encounters by ensuring that smoothing, rounding, and interpolation errors do not strongly impact the targeted data in the middle of the trajectory. Additionally, the higher standard for data quantities ensures all trajectories hold to a higher standard for data requirements.

Buffering the data works to limit data artifacts that could emerge in the following actions during these portions of the encounter generation process:

Data Processing Step: Horizontal and Vertical Smoothing and Track Interpolation.

Feature Measurement: Smoothing of True Airspeed, Acceleration, and Turn Rate.

To fully implement the encounter duration increase special consideration was given to the following steps to be consistant to the 160 second trajectory.

Initial Data Requirements: Data requirements were set assuming a 160 second trajectory.

Trimming to Capture Period: Capture period had to be set to the increased encounter duration.

Final Data Requirement: Data requirements were set assuming a 160 second trajectory.

Due to extending the processed trajectory the following additional steps were impacted, resolutions are detailed for addressing these uncaught consequences for future iterations.

Stationary Track Encounter Disqualification: This step in Data Processing evaluates the full trajectory, if no distance is traversed then a track is deemed stationary and the encounter is discarded. In the ECEM implementation this assesses the 160 second trajectory, which would enable movement in the buffer regions to prevent a track from being disqualified. Optimal implementation would evaluate only the 120 second trajectory period to ensure this step functions as designed.

Parallel Runway Encounter Disqualification: Disqualification for encounters if they occur too close to a runway is also performed on the full encounter. As is this check runs on the extended encounter duration which may result in encounters being disqualified due to the buffer region being within proximity of the airport although the targeted encounter duration would not merit disqualification. Optimal implementation of this step would involve assessing only the targeted encounter duration for proximity to airports with parallel runways.

Formation Flight Encounter Disqualification: Disqualification of encounters due to flying in formation assesses the full trajectory for similar flight maneuvers. As implemented, flights could be disqualified due to similar flight patterns in the buffer periods of the extended encounter. While identification of formation flight behaviors in the buffer periods may arguably classify the full encounter as a formation flight, discarding the full encounter is an extension of that assumption. Optimal implementation of this check would as originally targeted only evaluate the targeted encounter period.

6.4 INCREASED DATA REQUIREMENTS

Minimum data requirements were increased in two places in the Data Processing step. This implemented a higher threshold for quality and accuracy for an encounter to qualify to be used in the model. There were several reasons for this increase:

Artifact Reduction: The primary driver for increased data requirements was to reduce or eliminate data artifacts that were identified from the CEM evaluation. Detailed explanations about the artifacts are present in the Targeted Improvements section.

Improved Encounter Quality: A secondary reason for increasing the data requirements was to improve the overall accuracy and quality of the model. The relationships and distributions within the model are only as accurate as the observed encounters that are used to build the model, by increasing the data requirement for the encounters that serve as the building blocks for the model contents the model itself can more accurately extract and generate precise relationships and distributions. While higher resolution was critical for artifact reduction, increasing the minimum requirement for all encounters could be implemented due to the significant increase in identified encounters such that the requirement could be increased without imposing significant additional effort or delay to the model construction.

The two places data requirements were increased were in the Initial Data Requirement, (the second pass of the Initial Data Requirement), and the Final Data Requirement.

Initial Data Requirement: Data requirements were increased in two (of three) places in the data processing sequence. The initial raw data requirement for the CEM was ten data points over ten minutes centered on TCA. This was increased to 32 points of data over ten minutes centered on TCA for the ECEM.

The second pass of the data quality requirement occurs after the Track Overlap Trimming which can significantly reduce the duration of the encounter. The CEM would only accept an encounter of 50 seconds if it had ten data points within that period, or one data point every five seconds. The ECEM processing occurs on a 160 second encounter, so maintaining a proportional quantity of data requires 32 points.

These data requirement increases resulted in encounters that would at most require a scan rate of one every five seconds and at most lenient a scan rate of one every nineteen seconds which safely included the majority of scan rates observed in the data as is shown in Figure 12.

Final Data Requirement: The second data quantity increase occurs within the Final Data Requirement. The Final Data Requirement requires at least 155 data points for the ECEM after interpolation. The CEM required 45 data points for a targeted 50 second encounter duration. This could be either interpreted as up to give seconds short of the required duration or as 90% coverage of the encounter period. For this requirement the ECEM model requires the more conservative measure of 155 data points, rather than 90% of 160 seconds (144 data points).

6.5 ADDITIONAL CUT POINTS

A significant change to the model was the addition of new cut points to a number of variables. The addition of new cut points served the following purposes:

Prevent Markov Equilibrium: The primary motivation for increasing the number of cut points was to prevent or delay reaching a Markov Equilibrium when generating an encounter. Preventing a Markov Equilibrium only requires additional cut points in the Trajectory Generation dynamic variables: Acceleration, Vertical Rate, and Turn Rate. The added cut points for each variable in the Trajectory Generation step are shown in Table 8.

Distribution Mismatches: The secondary motivation for the addition of cut points was to more accurately generate distributions matching that of the observed encounters. The ECEM like the CEM utilizes uniform distributions within each bin as defined by the cut points. Fewer

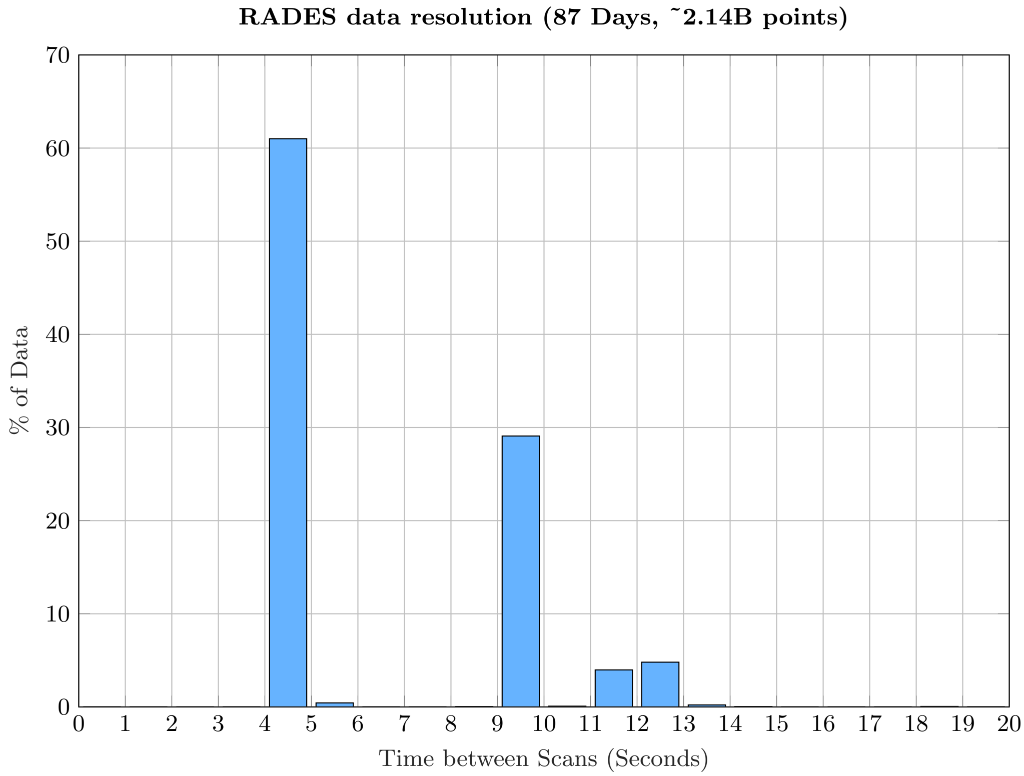

Figure 12. Scan rates of raw RADES data within the ECEM dates.

bins results in a rougher matching, a greater number of bins enables the distributions to more closely match the observed distributions. Additional details about Distribution Mismatches can be found in Targeted Improvements, Section 7.

Improved Encounter Quality: Additional improvements to encounter quality prompted the addition of cut points to other variables within the encounter model. Increasing the number of cut points for variables would enable the model to more precisely represent relationships between variables. This would result in generated encounters that more closely represent observed encounters. All additional cut points are shown in Table 9 for the Initialization Network and in Table 8 for the Trajectory Generation Network.

Additional cut points can be added directly and easily to the quantization scheme for each variable. However, there were potential other impacts to consider when increasing the number of cut points for variables.

Model Complexity: Each additional bin will need to have a probability of being chosen by every bin of the parent variables as well as probabilities for what values any children variables could take. Additional cut points to variables that have many connections will have a greater impact on the network than variables that have few connections. Also, cut points within

TABLE 8

Added Cut Points–Trajectory Generation Network

Variable	**New**\Old **Cut Points**	**Units**	**Parents**	**Children**
Altitude Layer (L):	**1000**, 3000, **5000**, 10,000, 18,000, 29,000, **40,000**	feet AGL feet MSL	0	6
True Airspeed (v_1, v_2):	100, **150**, 200, **250**, 300, **350**, 400, **450**, 500, **550**	knots	0, 0	1, 1
Acceleration ($\dot{v}_1$, $\dot{v}_2$):	±**4**, ±**3**, ±2, ±**1**, ±0.25	knots/second	5, 5	0, 0
Vertical Rate ($\dot{h}_1$, $\dot{h}_1$):	±5000, ±**4000**, ±3000, ±**2500**, ±2000, ±**1500**, ±1000, ±400	feet/minute	2, 2	1, 1
Turn Rate ($\dot{\psi}_1$, $\dot{\psi}_2$):	±**7**, ±6, ±5, ±3.5, ±**2**, ±1, ±0.25	degrees/second	2, 2	1, 1

Parent counts include the current variable, children counts exclude the current variable.

the ECEM are kept consistent per variable; therefore increasing cut points within Turn Rate impacts the Initialization Network and the Trajectory Generation Network. Execution of the ECEM is currently quite efficient, however, it is important to remember that increasing the number of bins increases the time required for data ingestion, the size of the stored version of the model, and the time it takes to generate an encounter.

Data Allocation: Adding a cut point splits the data for a bin into two bins, therefore it is important to consider the location and resulting distribution of data when adding cut points. While there will always be areas within the model where data is sparse, it is important to consider if addition of a cut point may degrade the model by spreading the available data too thinly across encounter bins.

Data Overfitting: Even with sufficient data within every bin, excessively narrow ranges within the bins will overly restrict the model. The bin ranges are a measure of trajectory variation since any value within a bin can be selected if that bin is chosen. Greater numbers of bins will necessitate breaking the range into smaller ranges and will decrease the inherent flexibility for generating more diverse trajectories. Therefore cut points must be added judiciously to prevent overfitting and constraining the model.

6.6 DYNAMIC ACCELERATION

The CEM model allowed only a constant acceleration over the course of a trajectory resulting in flights constantly accelerating, constantly decelerating or constantly maintaining speed. For a

TABLE 9

Added Cut Points–Initialization Network.

Variable	**New**\Old **Cut Points**	**Units**	**Parents**	**Children**
Altitude Layer (L):	**1000**, 3000, **5000**, 10,000, 18,000, 29,000, **40,000**	feet AGL feet MSL	0	11
True Airspeed (v_1, v_2):	100, **150**, 200, **250**, 300, **350**, 400, **450**, 500, **550**	knots	4, 3	4, 5
Acceleration ($\dot{v}_1$, $\dot{v}_2$):	±**4**, ±**3**, ±2, ±**1**, ±0.25	knots/second	3, 3	1, 1
Vertical Rate ($\dot{h}_1$, $\dot{h}_1$):	±5000, ±**4000**, ±3000, ±**2500**, ±2000, ±**1500**, ±1000, ±400	feet/minute	1,1	3, 3
Turn Rate ($\dot{\psi}_1$, $\dot{\psi}_2$):	±**7**, ±6, ±5, ±3.5, ±**2**, ±1, ±0.25	degrees/second	2, 3	1, 0
Horizontal Miss Distance: (hmd)	500, 0.5, 1, **1.5**, **2**, **2.5**	feet NM	4	1
Vertical Miss Distance: (vmd)	100, 200, ..., 900, **1000**, **1200**, **1400**, **...**, **5800**	feet	3	1

50-second encounter this was a reasonable assumption but it is significantly less realistic for a 120 second duration encounter and as a result it was determined that the acceleration of the aircraft must be variable.

Several options were considered including:

- Adding speed as a dynamic variable to the Trajectory Generation Network, leaving acceleration in the Initialization Network but not using the values selected.
- Adding acceleration (as an addition to speed) to the Trajectory Generation Network and substituting the values selected for the acceleration that was added to the prior speed each time step as in the CEM.
- Adding acceleration (as a multiplier speed) to the Trajectory Generation Network and multiplying the speed by the new acceleration values each time step.

Modeling acceleration as an addition to speed was decided on as the framework already existed in the model, though it was unused. This also required the minimum amount of modification to the model, acceleration needed to be added into the transition matrix similar to Turn Rate and Vertical Rate. Figure 13 shows the ECEM Trajectory Generation Network with acceleration included.

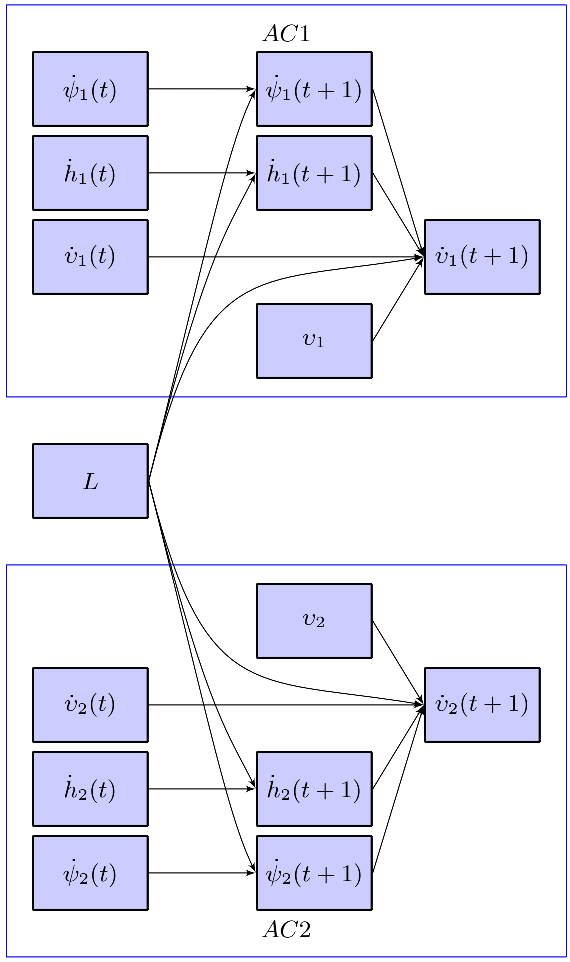

Figure 13. The Trajectory Generation Network with acceleration as a dynamic variable.

6.7 INCREASED DATA FOUNDATION

The number of observations contributing to the model was also significantly increased to a baseline of five million encounters. Multiple factors contributed to this decision:

Additional Cut Points: The significant increase of cut points across multiple variables resulted in the baseline data being spread and split amongst a greater number of bins than previously. To ensure sufficient data within bins to maintain encounter quality and to have sufficient data to have diversity within the bin values as well required a significant increase for the number of encounters that would build the model.

Sparse Data Areas: Additionally targeted was to increase the baseline number of encounters to increase the numbers of rare encounters that contributed to the model. During the evaluation of the extended CEM one potential improvement identified was to increase the counts in bins that had only few data points to work from. These bins were frequently at the edges of the variable ranges and represented encounters with more extreme maneuvers. While the model would produce encounters based on the available values, greater numbers of encounters contributing to these 'low data areas' would provide greater confidence in the legitimacy of encounters generated from those areas.

Greater Encounter Diversity: The addition of many new encounters was also desired to increase the modeling capability to areas previously without sufficient radar coverage. A higher baseline data requirement would increase the likelihood for the inclusion of new encounters it would be possible to expand the encounter generation space enabling generation of a broader range of encounters.

The baseline data goal was achieved, requiring 90 days of the RADES data and resulting in 5,530,665 encounters contributing to the model.

7. TARGETED IMPROVEMENTS

7.1 ARTIFACTS

During the evaluation of the CEM data artifacts were identified in the processed trajectories used to build the model. These artifacts are in the form of greater numbers of extreme values at the beginning and end of processed trajectories. These are evident in Figure 14 and 15 which show the range of Turn Rates across the encounter duration for 50 second encounters and 120 second encounters respectively from the 2008 Encounter Archive. The distinctive "bow-tie" shown in these graphs indicate a greater spread of turn rates at the tails of the encounters. Turn Rate specifically had significant artifacts; artifacts were not sufficiently present in Vertical Rate to warrant concern.

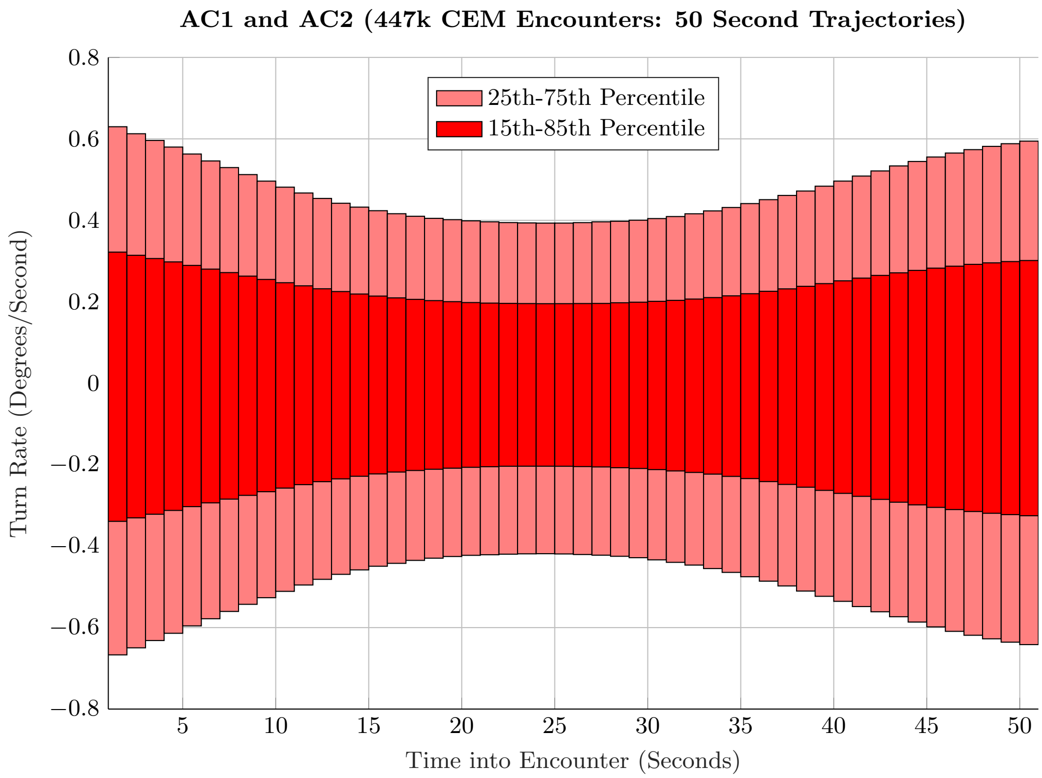

Figure 14. A distinctive "Bow Tie" shape reveals the presence of data artifacts in the processed CEM trajectories.

Once identified, these patterns were strongly suspected to be the result of rounding due to a few key observations:

Artifact Duration: In both the 50 second encounters as well as the 120 second encounters it was noted that the duration of the "flares" of the bow-tie were consistently 20 seconds. This pattern is difficult to discern in Figure 14 since the beginning and trailing artifacts almost

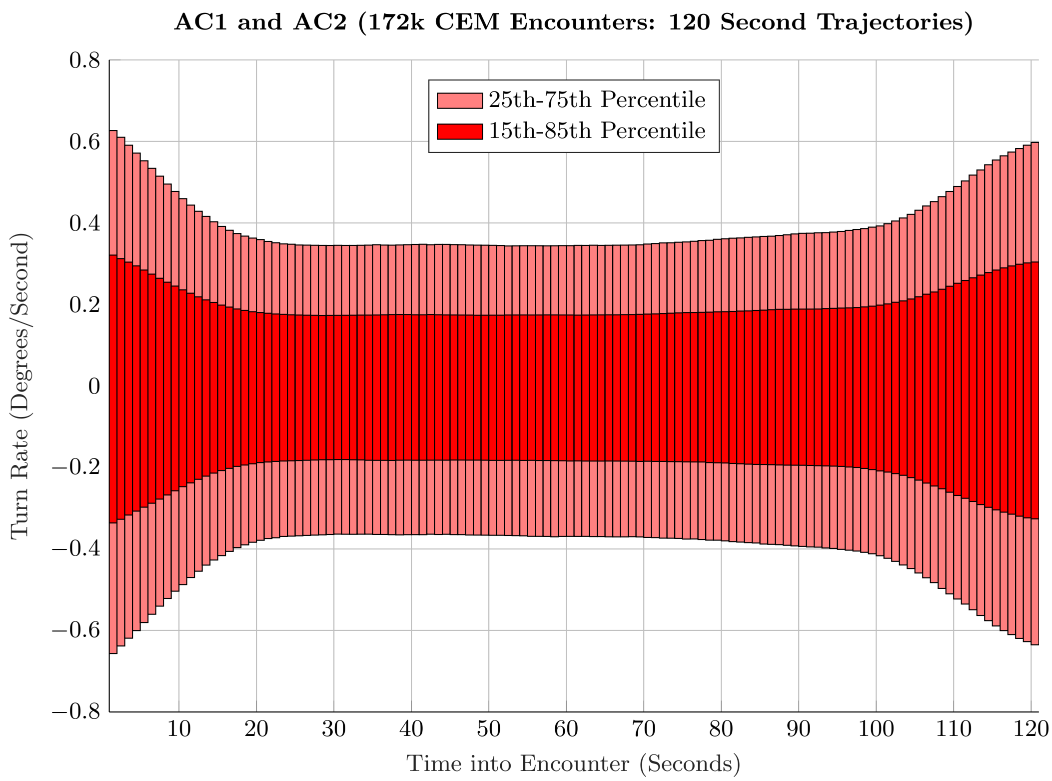

Figure 15. 120 second trajectories from the 2008 Encounter Archive also gain a "Bow Tie" artifact pattern after processing.

comprise the full duration of the encounter, but there is a short period from ten seconds to 30 seconds where the spread stabilizes briefly. It is clearly observed in Figure 15 where after 20 seconds into the encounter the shape of the Turn Rate spread becomes nearly constant until 100 seconds in where the spread of Turn Rates increases again.

Artifact Pattern: If believed, the pattern shown by the artifacts would imply that there were a significant number of encounters that would be engaging in turns only at the tails of the observed encounter period. Given that this model studies correlated encounters it would not be unexpected to see a turn begin immediately preceding TCA for the encounters, but significantly more turning occurring at the beginning of the capture period had no logical basis

Pattern Inconsistency: Finally, it was observed that since the set of 120 second encounters is a subset of the 50 second encounters and these values were all accurate then the 50 second "bow-tie" should be visibly present within the 120 second data. However, beyond the 20 second increased spreads at the tails of the encounter the rest of the Turn Rate distributions are unchanging.

These artifacts caused a number of problematic results:

Disproportionate Extreme Value Prevalence: The extreme values caused by the artifacts are ingested into the encounter model, resulting in distributions that are misrepresenting the actual makeup of variable values. Particularly significant artifacts may even populate previously empty bins allowing generation of more extreme encounters although the probability of those encounters being generated may still be low.

Initialization Network Distribution: Due to the artifacts occurring at the tails of the trajectories in combination with the Initialization Network selecting the Trajectory Generation variables at starting time the impact on the model and the generated trajectories would be considerable. The initialization variable distributions were selected exactly where the artifacts were most prevalent and severe.

Trajectory Generation Network Distributions: The "memoryless" structure of the Trajectory Generation Network results in all of the dynamic variables in the network being condensed regardless of time into a single set of distributions. This enables the probability of selecting an extreme value to be dispersed due to the lack of artifacts in the middle of the encounter; however it also enables extreme values to be selected at any time when generating an encounter.

Validation Discrepancies: Conditions at and immediately around TCA are some of the most critical to have accurately representing observed encounters. However, TCA in the ECEM occurs ten seconds from the end of the encounter which is well within the 20 second period at the end of the trajectory that the artifacts occur within. Turn Rate, Vertical Rate, and Acceleration are propagated from the beginning of the encounter and so differ from the distributions at TCA however, validation against the processed observed trajectories guarantees a significant difference due to the presence of the artifacts. Other validation metrics that involve or assess Turn Rate, Vertical Rate, and Acceleration will similarly have discrepancies due to the artifacts. Figure 16 shows the clear presence of artifacts in the observed data. Figure 16 shows the percentage of tracks that are not turning as a function of time through the encounter. The low frequency of straight flight observed at the beginning and end of the encounter is due to the artifacts and creates an impossible and inaccurate expectation when compared to the generated encounters.

Several changes were implemented to reduce/remove the artifacts from the ECEM. Details on the exact implementation of these changes are present in Model Changes and Updates, Section 6.

Buffered Encounter Duration for Processing: Since the artifacts were observed to only occur within 20 seconds of the tails of the encounter period, the first direct change to reduce them was to extend the encounter duration during processing by an additional 20 seconds at both ends.

Increased Data Requirements: An additional step was taken to ensure throughout the Data Processing steps that the trajectory information would be evaluated and assessed to

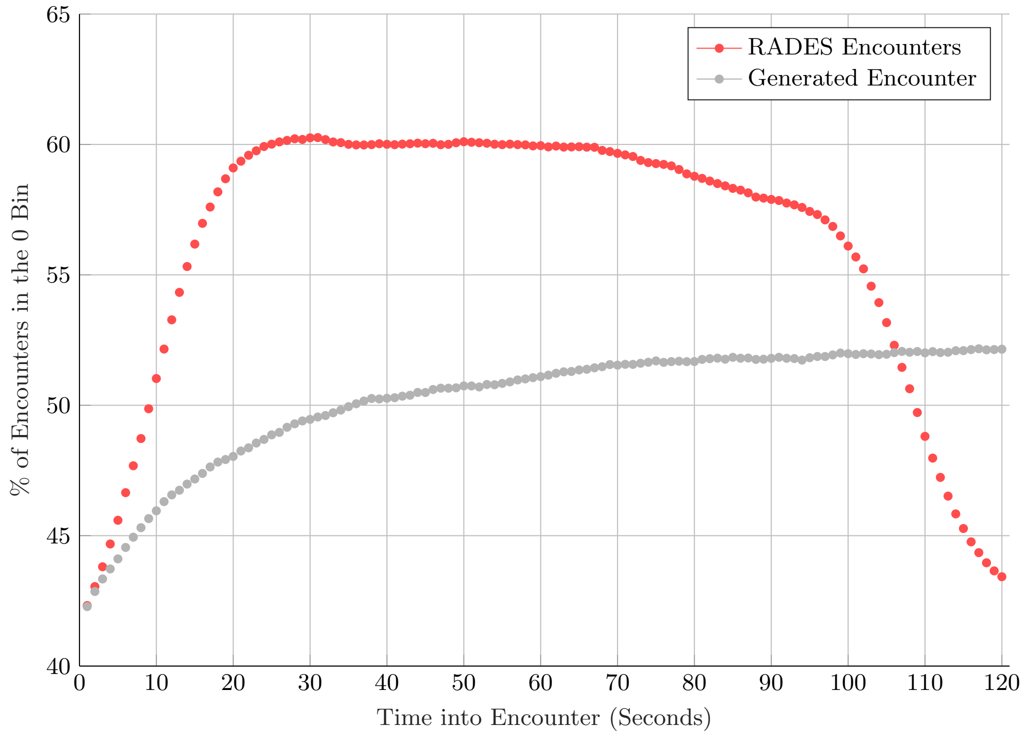

Figure 16. The percentage tracks flying straight both observed (with artifact) and generated.

a higher set of requirements. Increased amounts of data, particularly near the tails, would decrease or prevent extreme values caused by smoothing.

All dynamic variables in the ECEM were evaluated to measure the impact of the changes to resolve the data artifacts.

Turn Rate: Turn Rate artifacts were greatly reduced. Figure 17 shows the spread of Turn Rates across time for ECEM trajectories. No “Bow Tie” pattern is evident. Figure 18 shows the percentage of encounter aircraft in the “zero” Bin for Turn Rate which are stable across the encounter duration.

Vertical Rate: Artifacts in Vertical Rate were insignificant prior to the implemented changes, and continued to be insignificant after the changes.

Acceleration: Acceleration was not a dynamic variable prior to the ECEM, however significant artifacts were found in Acceleration despite the implemented changes. In Figure 19 one-third of the ECEM AC1 trajectories clearly show the “Bow Tie” present in Acceleration at the encounter tails similar to the artifacts originally observed in Turn Rates.

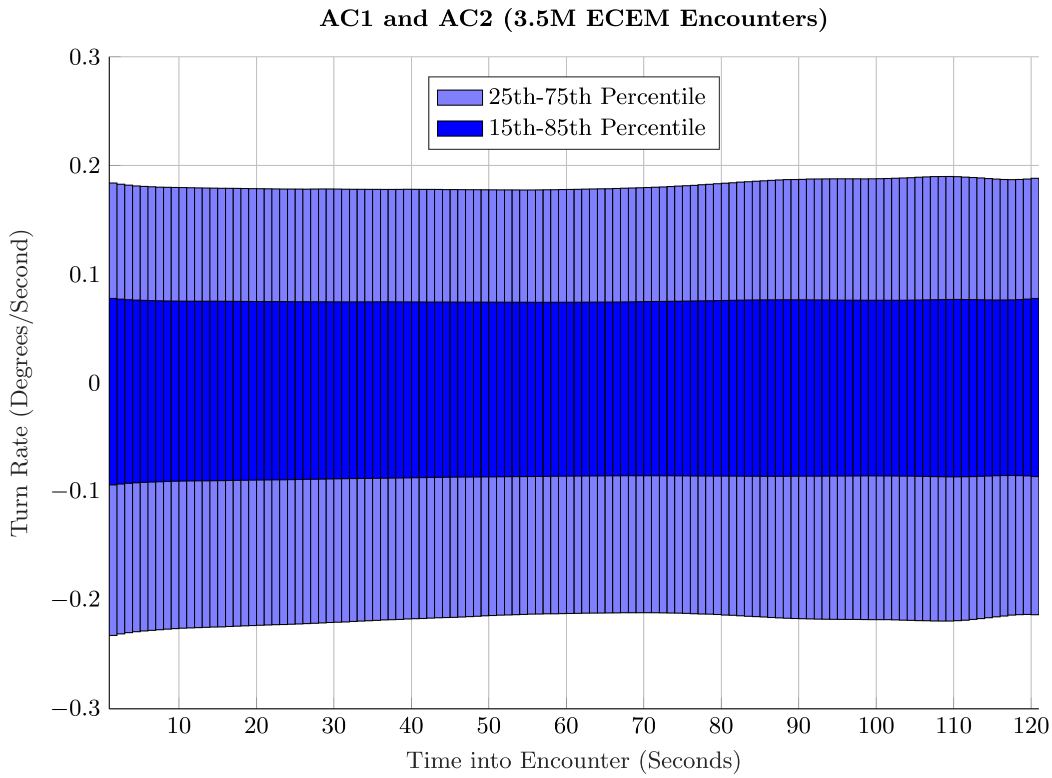

Figure 17. After implementation of multiple resolving changes, Turn Rate artifacts are significantly reduced in the ECEM.

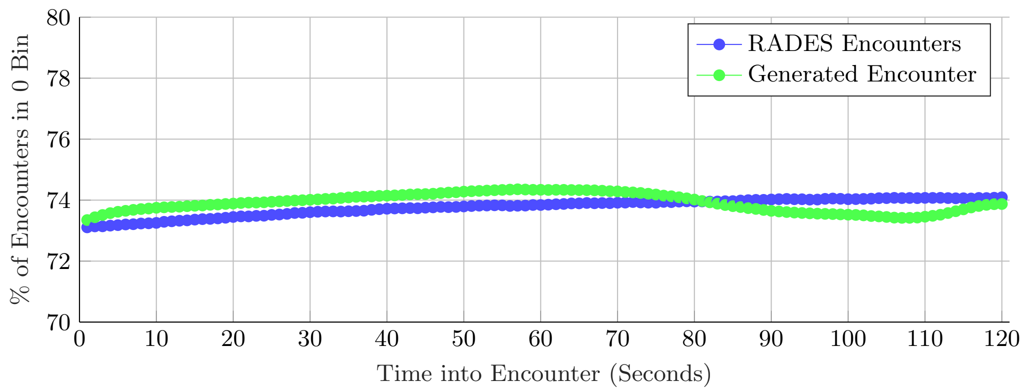

Figure 18. The percentage tracks flying straight both observed and generated for the ECEM.

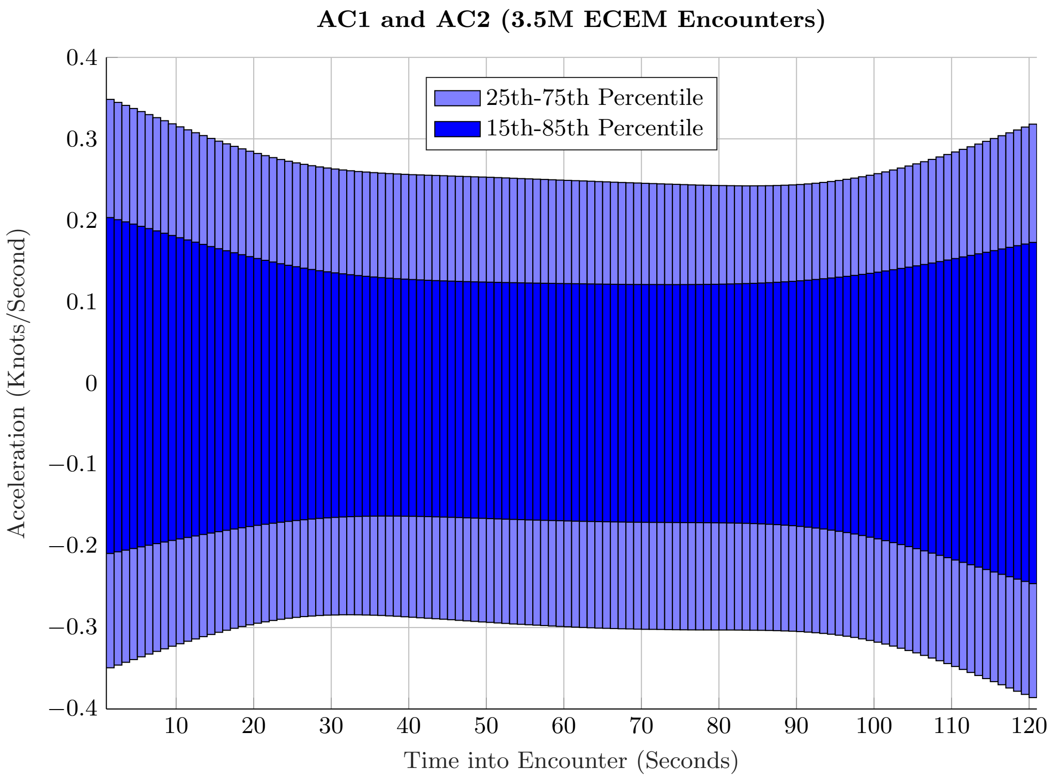

Figure 19. Despite implementation of multiple resolving changes, acceleration artifacts are detected.

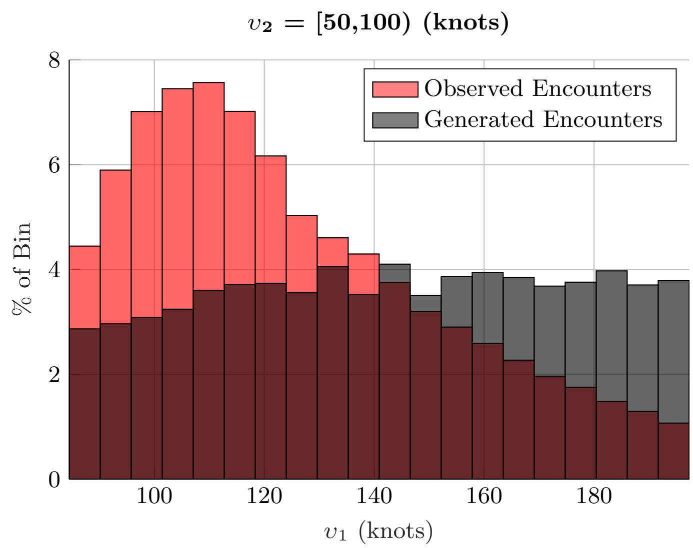

Figure 20. The distribution of v_1 for observed and generated encounters in the CEM where v_2 is within [50,100) knots, AC1 is a discrete code aircraft, and the Altitude Layer is 1: [1k, 3k) feet.

7.2 DISTRIBUTION MISMATCHES

Another improvement identified when assessing the CEM during development of the ECEM was the rough matching between generated distributions and observed distributions. Figure 20 highlights this aspect of the CEM where speed for AC1 is represented by a single uniform distribution between [100,200) knots. Quite a different distribution is present in the observed speeds for AC1 however, which shows that speeds between 100 and 120 knots are the most prevalent and faster and slower speeds than those are of decreasing frequency.

The simple solution to lessen these mismatches was to add additional cut points to increase number of bins that would make up variable ranges. Additional information on how and where cut points were added can be found in the Model Changes and Updates section. The additional cut points would increase accuracy of distribution mismatches in a two key ways:

Increased Numbers of Distributions: Each value for a parent variable will have a probability for each of the child variable's values. The more bins available in the parent variable the more unique distributions there will for the child variable. An example from ECEM is the increase of Altitude Layers from five to eight; subsequently Airspace Class has more distributions one specific to each of the eight layers. However, Airspace Class still only has four possible values. Therefore the parent variable increased the number of cut points, creating a set of probabilities for each Airspace Class per Altitude Layer; however the number of possible values for Airspace Class (the child variable) remains the same.

Increased Distribution Resolution: The more bins in the child variable however, the more precisely the distributions determining a child variable's value can be represented and leveraged. In the ECEM Vertical Rate for AC2 is a child variable of Altitude Layer. By increasing the number of Altitude Layers from five to eight the number of distributions for Vertical Rate increases (one for each Altitude Layer). The precision of those distributions is also be increased. Vertical Rate used to be represented by eleven bins, so each distribution would be comprised of eleven probabilities (adding up to one), one value for the likelihood of the Vertical Rate number to fall in that bin given the Altitude Layer used. In the ECEM six additional cut points were added, therefore seventeen probabilities make up each distribution. Another way to understand this change is by understanding that in six of the old bins, the new model now will be able to assert whether the value should be selected from the lower half of the bin or the upper half of the bin.

The main observations of the resulting ECEM data after cut points had been added and distribution matching assessed were:

Sufficient Data Enabled Improved Matches: All bins with sufficient amounts of observed data to produce a solid distribution shape were better matched by the generated data using additional cut points. Figure 21 shows the distribution in ECEM that occupies the same characteristic space as the distribution in Figure 20. In Figure 21 the distribution resolution was increased due to additional cut points in the child variable. This is apparent as the percentage increase in generated encounters between 100 and 150 knots for AC1 to better match the observed distribution, and then the lower percentage rate of generated encounters from 150 to 200 knots. (In Figure 20 there could only be a single uniform distribution from 100 to 200 knots.)

Numerous Low Data Bins Created: The addition of cut points did however split bins where only a few data points went to one of the new created bins. While encounters could and were generated for some of these bins, insufficient quantities of data do not produce a distribution that can be well matched or is necessarily sensible to attempt to match.

Numerous Sub-distributions Still Present: Despite the addition of many cut points to some variables, matching the distributions for some variables would require even more additional cut points. Even within Figure 21 the generated distribution is "blocky" compared to the observed distributions, to better match the observed data cut points would need to be added and/or shifted to improve the match.

Additonal distribution graphs are available in Appendix F.

7.3 LOW DATA AREAS

Low data areas are bins in which there are few data points, which raise the concern that there may not be sufficient data to accurately represent encounters of those specific characteristics. Within the encounter characteristic space there are significant areas to which many observed encounter contributed to, encounters generated from those areas are well backed by a significant foundation of data. Areas within the model that have numerous data points validate themselves, many observations of similar behavior provide evidence that the relationship observed is not an

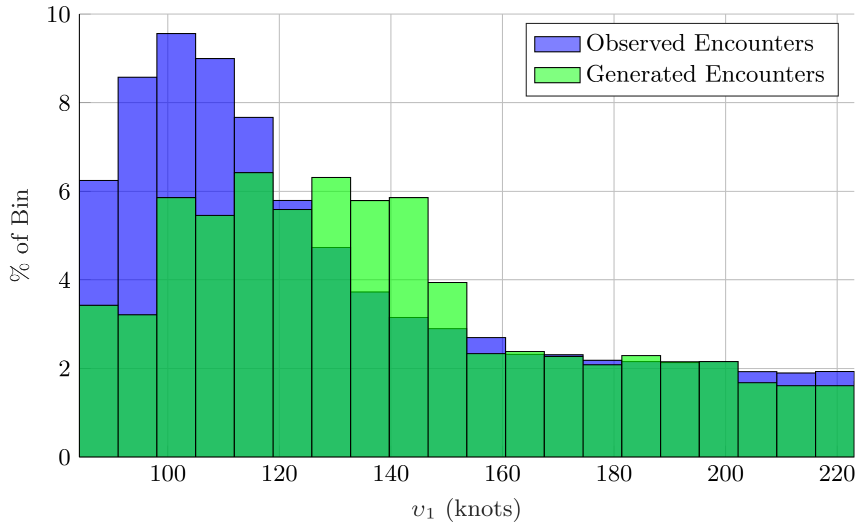

Figure 21. The distribution of v_1 for observed and generated encounters in the ECEM where v_2 is within [50,100) knots, AC1 is a discrete code aircraft, and the Altitude Layer is 2: [1k, 3k) feet.

outlier. However, there are areas of the model where only few encounters matched the characteristics; these areas often do not have sufficient data to form a smooth distribution. If there is a single data point present in a bin, the ECEM will create encounters based on it, this becomes problematic when using the ECEM to generate specific encounters. When using the ECEM without constraints the occurrence of these rare encounters would have low probabilities based on the low counts of data available in them, however if the ECEM is used to generate specific types of encounters these rare characteristic combinations could be produced exclusively—but still based on very few data points.

To resolve Low Data Areas three mitigating strategies were identified:

Increased Data Foundation: The perfect solution would be to gather sufficient data for all types of encounters used in the model. One effort to attain this goal was through significantly increasing the number of encounter used in the model. Additional details of the data foundation increase are laid out in Model Changes and Updates, Section 6.

Bin Merging: Another option was to take bins with low data counts and merge them with adjacent bins. Bins are defined by the cut points; therefore values slightly past the cut point threshold could segregate a small amount of encounters, creating a low data area. While sound in theory, the implementation of the ECEM is such that cut points must be consistent for a single variable and a cut point that would merge a low data bin in one instance would additionally remove that cut point where it successfully split bins with more than sufficient data in others. For this reason, bin merging could not be precisely applied and was not further pursued as a potential solution.

Encounter Rejection: A final solution identified was to reject encounters generated from a low data bin. This is a sub-optimal solution as it would require evaluation of every generated encounter, a prior assessment of what encounter characteristics would qualify for disqualification, and would result in discarding generated encounters reducing generation efficiency. This remains as a potential mitigation solution however as mentioned a methodology for disqualification would have to be identified and applied.

8. VALIDATION

Validation for the ECEM was structured to answer two main questions:

Did the ECEM work properly: do the generated encounters match the observed encounters? This section answers this question.

Were the targeted improvements addressed: were applied changes successful in improving the model? This is detailed in the Targeted Improvements section.

8.1 METHODOLOGY

Validation of the model was performed in two ways:

Statistical Measurement: This part of the model validation is quantitative. The Bayesian network for the ECEM is compared to a Bayesian network composed of encounters generated from the ECEM.

Visual Assessment: This part of the model validation is qualitative. Distributions for both the ECEM and the encounters generated by the ECEM are graphed together and visually compared.

8.1.1 Statistical Measurement

This process of measures the accuracy and functioning of the ECEM mathematically and is performed by generating a new model using the synthetic encounters and then directly comparing the variable bin counts to the original ECEM.

Step 1. Generation: One million encounters were generated using the ECEM. Provided the ECEM was created and used for generation properly, generated encounters should be created according to the probabilities stored in the ECEM. Encounters are generated using a simple generation script designed for the CEM.

Step 2. Ingestion: The generated encounters were subsequently formatted to the data structure needed for ingestion into the model. The Bayesian network structure used is exactly the same as the ECEM, including identical cut points and zero bins. The same code used for generating the ECEM from the observed data is used as well to minimize any differences in processing.

Step 3. Comparison: Comparison of the two Bayesian networks directly to each other would be inaccurate since a different number of encounters were used to build each. However, the percent of traffic in each bin for a variable can be used as an equal measure to span the difference. This is not a perfect measure, evaluation of a variable will to incorporate discrepancies from the parent variables. Additionally, due to a lesser number of generated encounters there will very likely be small differences as rare bins may not have any encounters generated within them. As an overall assessment of accuracy this method does provide a single comprehensive number for each variable for how well the two networks match.

8.1.2 Visual Assessment

The qualitative measurement for assessing the ECEM is simple but enables identification of subtle differences and enables evaluation of each distribution. The Visual Assessment is straightforward:

1. Generate one million encounters using the ECEM.
2. Create graphs showing observed and generated encounter distributions.
3. Visually inspect distributions for significant differences.

Step 1. Generation: It is necessary to have a comparison set of ECEM generated data for the Visual Assessment. Prior evaluation has shown that one million encounters are sufficient for evaluation however more if possible is better. Utilizing the same set of generated encounters for the Visual Assessment as was used for the Statistical Measurement is sufficient.

Step 2. Creation: To perform the Visual Assessment efficiently each distribution is individually graphed according to the percentage of data in set intervals for both the generated and observed data. For ease of analysis these graphs are then overlaid on each other. Not all variables are easily evaluated in this manner due to the number of distributions a variable can have, but the effort of performing this assessment is matched by the depth of detail it reveals about the model's accuracy.

Step 3. Inspection: Inspection aims to assess multiple aspects including: how well the ECEM distributions matched the data, how well the existing cut point placement enables the model to match the data, what if any underlying sub-distributions exist in the data and could the model better conform to them?

The Visual Assessment was performed using a regular interval bar graph which enables the following precise evaluations:

Cut Point Assessment: This evaluation will reveal exactly the distribution of encounters within bins providing a measurement of distribution accuracy. The Visual Assessment complements that evaluation by evaluating distributions using regular intervals across the variable range. This enables evaluation of how well the cut points and distributions match the observed data.

Sub-distribution Identification: High resolution distributions using regular intervals enables detailed graphs of the frequency of values for observed versus generated data.

To accurately evaluate the data in this format it is important to recognize aspects of the encounter structure that will result in prominent and different distributions than the observed encounters. These include:

Zero-bins and Value Limits: Zero bins convert all values in a range around zero to zero, often resulting in a single bar surrounded by a lack of data. Similarly, value limits prevent the generation of values outside a specified range for instance *vmd*'s maximum value is 6000 feet. Figure 22 shows an example of the presence of a Zero Bin, and Figure 23 shows an example of the value limits for *vmd*.

Encounter Count Discrepancy: Particularly for distributions with a smaller number of observed counts, even generating one million encounters may not be sufficient to match the distribution or even produce an encounter of those specific characteristics. Figure 24 shows an likely example of when distributions do not match due to insufficient numbers of generated encounters.

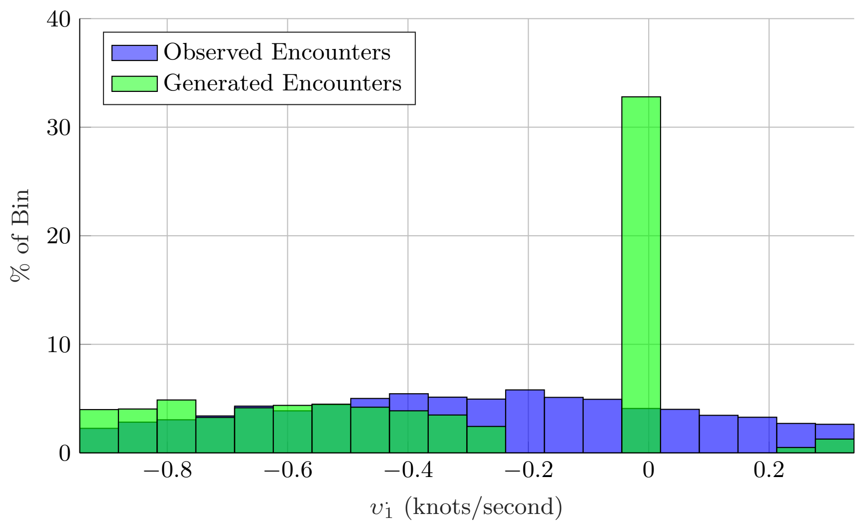

Figure 22. The distribution of $\dot{v}_1$ for observed and generated encounters where L=4, $\dot{h}_1$=[-400,400), and v_1=[350,400).

8.2 RESULTS

8.2.1 Statistical Measurement

Each distribution from the ECEM was compared to the same distribution present in a matching Bayesian Network populated by one million encounters generated by the ECEM. To enable equivalent comparison each bin in each distribution was normalized to the percent of the total within each bin. The number used as a metric of match is the minimum of the observed and generated bin percentages. The minimum percentage of observed and generated is the overlap for the two networks; the sum of the minimum percentages across all bins produces the total match for a variable. Table 10 and Table 11 show the percentage difference for each of the network variables. Figure 25 shows an example of the bin percentages for χ.

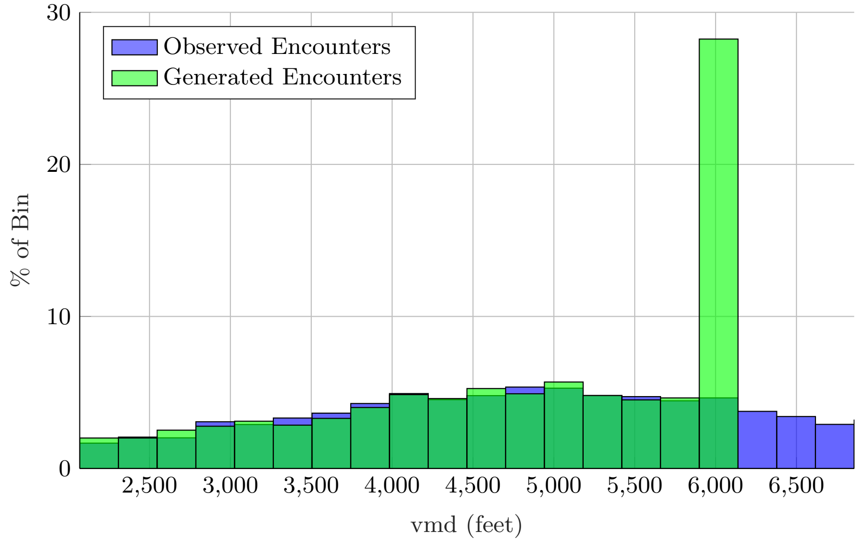

Figure 23. The distribution of vmd for observed and generated encounters where $\dot{h}_1$=[-400,400), $\dot{h}_2$=[1500,2000), and L=4.

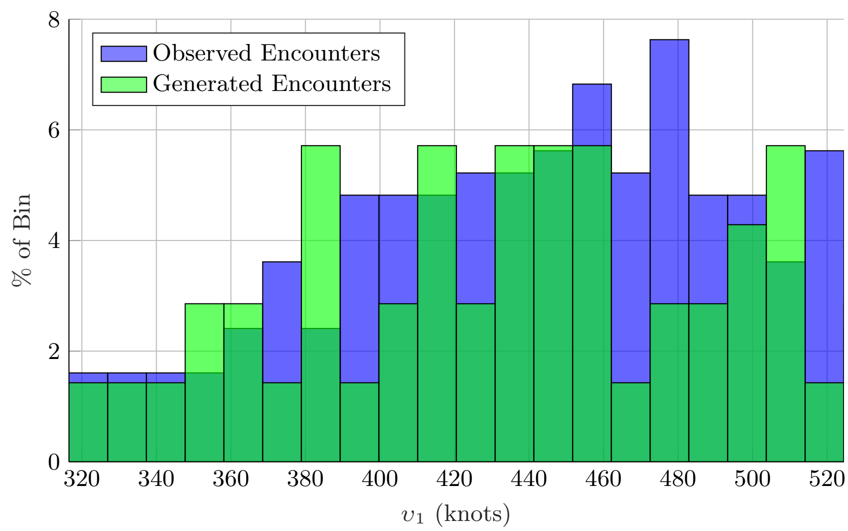

Figure 24. The distribution of v_1 where L=7, h_1=[-1500,-1000) feet/minute, v_2=[250,300) knots, and AC1 is a discrete code aircraft.

TABLE 10

Statistical Measurement: Initialization Network

Variable	Match %	Parent Variables	Total Bins
Airspace Class (A):	99.67	L	32
Altitude Layer (L):	99.74		8
Bearing (χ):	89.98	β, hmd	168
Approach Angle (β):	95.65	v_1, v_2	1452
Aircraft Category AC1 (C_1):	99.64	L, A	64
Aircraft Category AC2 (C_2):	99.56	L, A	128
True Airspeed (v_1):	84.42	L, C_1, $\dot{h}_1$, v_2	32912
True Airspeed (v_2):	84.73	L, C_2, $\dot{h}_2$	2992
Acceleration ($\dot{v}_1$):	95.29	L, $\dot{h}_1$, v_1	16456
Acceleration ($\dot{v}_2$):	89.90	L, $\dot{h}_2$, v_2	16456
Vertical Rate ($\dot{h}_1$):	99.25	L,	136
Vertical Rate ($\dot{h}_2$):	99.36	L,	136
Turn Rate ($\dot{\psi}_1$):	98.33	v_1, $\dot{v}_1$	1815
Turn Rate ($\dot{\psi}_2$):	93.93	v_2, $\dot{v}_2$	27225
Horizontal Miss Distance (hmd):	68.62	L, vmd, v_1, v_2	243936
Vertical Miss Distance (vmd):	81.34	L, $\dot{h}_1$, $\dot{h}_2$	83232

TABLE 11

Statistical Measurement: Trajectory Generation Network

Variable	Match %	Parent Variables	total Bins
Vertical Rate ($\dot{h}_1(t+1)$):	98.46	L, $\dot{h}_1(t)$	2312
Vertical Rate ($\dot{h}_2(t+1)$):	98.71	L, $\dot{h}_2(t)$	2312
Turn Rate ($\dot{\psi}_1(t+1)$):	95.15	L, $\dot{\psi}_1(t)$	30600
Turn Rate ($\dot{\psi}_2(t+1)$):	96.76	L, $\dot{\psi}_2(t)$	30600
Acceleration ($\dot{v}_1(t+1)$):	79.19	L, $\dot{v}_1(t)$, $\dot{h}_1(t+1)$, $\dot{\psi}_1(t+1)$, v_1	2715240
Acceleration ($\dot{v}_2(t+1)$):	76.31	L, $\dot{v}_2(t)$ $\dot{h}_2(t+1)$, $\dot{\psi}_2(t+1)$, v_2	2715240

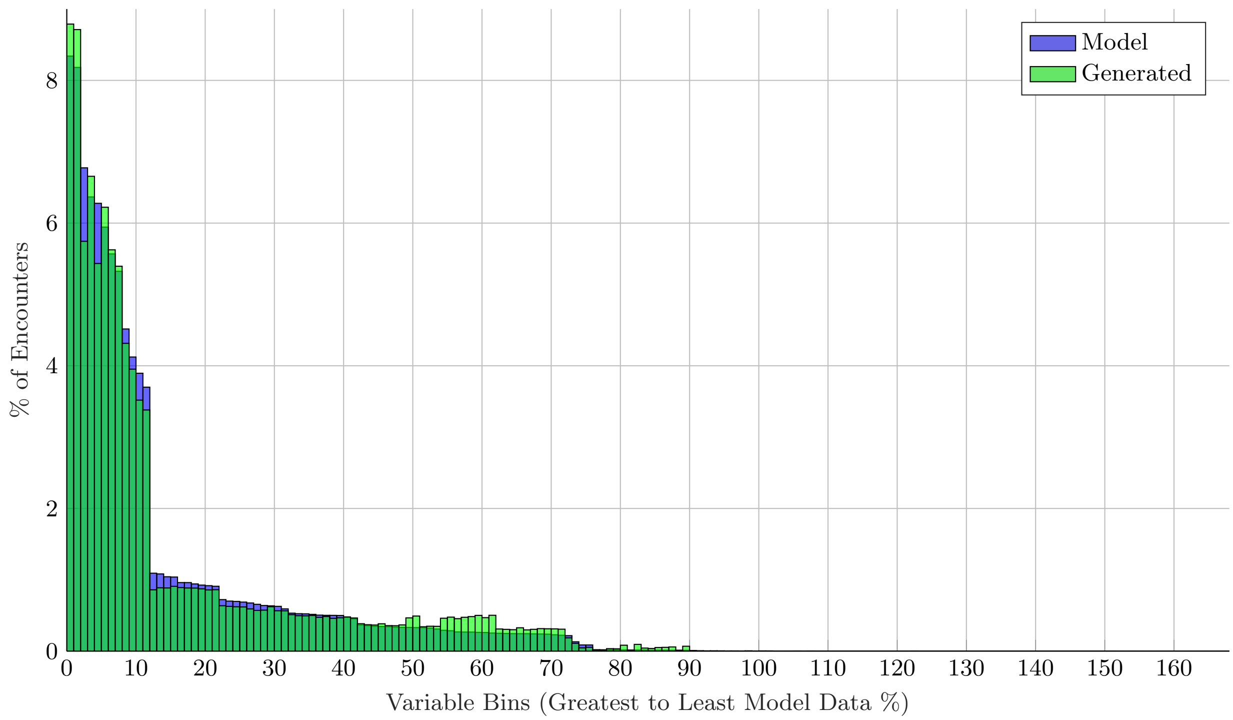

Figure 25. The percentage a bin in χ was used for observed and generated encounters. (Bins ordered from greatest percentage occuring in the model to least.)

9. POTENTIAL FUTURE IMPROVEMENTS

Future iterations of this or other encounter models may wish to consider the following ideas for model improvements.

9.1 HIGHER QUALITY DATA

During the course of development several potential data improvements were identified including:

- **DTED Level 1**: A higher resolution than the DTED level 0 (900 meter) resolution data used for this model, DTED Level 1 with 90 meter of the surface would enable more accurate estimates of AGL for the encounters.
- **SUA Use Times**: This model eliminates all encounters where the TCA occurs in SUA. However, different encounter geometries may be lost due to this elimination criteria. Inclusion of SUA use times would enable more precise reductions of encounters that were likely military related, while enabling use of encounters that pass within SUA geographic boundaries when the particular block of SUA is inactive.
- **Radar Data Fusion**: A significant filter for usable encounters was the amount of data required to create a clean and detailed track. Fusion of data from multiple radars would greatly reduce the number of encounters discarded and would potentially greatly improve the quality of tracks.
- **Aircraft Type Information**: Considered in development but eventually discarded was the potential for additionally leveraging categories for aircraft (such as weight class) to better model the physical limitations, maneuvers, and characteristics by the aircraft itself. Unfortunately RADES data does not contain information about the aircraft model so this improvement was considered to be too significant an effort for this encounter model. Aircraft type could be obtained in future iterations by merging radar data with flight plan data.

9.2 IMPROVED DATA CLEANING

Retrospectively, improvements were identified that would produce better encounters in the Data Processing steps. These included:

- **Special Beacon Code Handling**: Aircraft are currently categorized as either discrete code aircraft or 1200 code aircraft. There exist a number of beacon codes that additionally have special meanings including: 7500 - hijacking, 7600 - radio failure, and 7700 - emergency. The behavior of aircraft squawking these codes are unlikely to be representative of the encounters desired. However rare these are likely to be, it would be good practice to check and remove any flights with specialty beacon codes from the data set.

- **Targeted Assessment of Encounter Duration**: This study expanded the encounter duration during processing to reduce data artifacts. As implemented this resulted in the potential for encounters to be discarded due to the errors found in the buffer period outside the actual encounter duration. While the buffers help considerably, other rejection tests must be fine-tuned to only disqualify encounters for criteria that occurred during the utilized encounter period.

- **Additional Acceleration Smoothing**: Data artifacts were targeted for reduction or removal in the ECEM, however artifacts were present in Acceleration during the final evaluations of the raw data. Future iterations of encounter models should seek to improve processing such that Acceleration artifacts are also removed or reduced if possible.

9.3 INCREASED ENCOUNTER DETAIL

As encounter durations increase, certain assumptions decrease in legitimacy. Future encounter models may need to consider a more detailed and flexible method of realistically representing observed encounters. The inclusion of dynamic acceleration in the ECEM was one such upgrade, made necessary by the failure of prior acceleration modeling to realistically stretch to a 120 second duration. Other potential considerations include:

- **Dynamic Airspace Class**: Currently Airspace Class is measured at TCA; however this is no guarantee that the rest of the encounter occurs in that same airspace class. Particularly for Airspace Class associated with an airport, encounters would be expected to transition in or out of the Airspace surrounding the airport and the behaviors of the encounters would likewise be expected to change.

- **Airspace Sub-Classes** Airspace class could additionally be divided into finer bins this could including classifying Class B airspaces as those that serve a "cornerpost" airport versus serving a "multiplex."

- **Dynamic Altitude Layer**: Altitude Layer is already slightly problematic for encounters such as those in the ECEM. Encounters with tracks that have a high rate of descent or ascent frequently change Altitude Layers. Although inclusion of Altitude Layer as variable will require that generated trajectories have a specific altitude to capture changes between layers reducing the flexibility of a model such as the ECEM. However, it is also important to keep the aircraft's expected maneuvers synonymous with the behavior for their altitude layer.

- **Beacon Code Handling**: Another simplifying assumption present in the ECEM is to reduce an aircraft down to either a 1200 code or discrete code aircraft. As more encounter types are tracked and desired to be modeled it may become necessary to distinguish within an encounter if a flight transitions from a 1200 code to a discrete code or vice versa.

9.4 PRECISE ENCOUNTER GENERATION

In the encounter generation process a single improvement was identified.

- **Dirichlets of Zeros**: The legacy generation method from the CEM uses a Dirichlet prior of ones as the foundation for the bin counts and calculating the bin probabilities. This results in non-zero probabilities for bins where no encounters have been observed with those criteria. While the probability of generating one of these non-existent encounters is low, it could be easily ensured that generated encounters have characteristics seen in the observed encounters by instead replacing the Dirichlet prior with one of zeros.

This page intentionally left blank.

A. MODEL PARAMETERS

This section describes all information available for the model parameters or network variables used in the ECEM. All constant variables used in the Trajectory Generation Network have the same properties as those same variables in the Initialization Network.

TABLE A.1

Details of Airspace Class (A)

Characteristic	Variable Values
Defined Time/Period	TCA
Used Time/Period	Full Encounter Duration
Bounds	N/A
Cut Points	N/A
Zero Bin	N/A
Bin Count	4
Parent Variables	L
Children Variables	C_1, C_2
Total Bin Count	32 (Table 5)
Value Equivalents	1 = B
	2 = C
	3 = D
	4 = O

TABLE A.2

Details of Altitude Layer (L)

Characteristic	Variable Values
Defined Time/Period	TCA
Used Time/Period	Full Encounter Duration
Bounds	N/A
Cut Points	N/A
Zero Bin	N/A
Bin Count	8
Parent Variables	None
Children Variables	$A, v_1, \dot{v}_1, \dot{h}_1, c_1, vmd, hmd, v_2, \dot{v}_2, \dot{h}_2, c_2$
Total Bin Count	8 (Table A.3)
Value Equivalents	1 = < 1,000 AGL
(Figure 3)	2 = 1,000 AGL–3,000 AGL
	3 = 3,000 AGL–5,000 MSL
	4 = 5,000 MSL–10,000 MSL
	5 = 10,000 MSL–18,000 MSL
	6 = 18,000 MSL–29,000 MSL
	7 = 29,000 MSL–40,000 MSL
	8 = $\geq$ 40,000 MSL

TABLE A.3

Bin probabilities for Altitude Layer

L	Bin Counts
1	20069
2	460354
3	680549
4	1220985
5	1453888
6	673961
7	963531
8	57328

TABLE A.4

Details of Aircraft Category for AC1 (C_1)

Characteristic	Variable Values
Defined Time/Period	Full Encounter Duration
Used Time/Period	Full Encounter Duration
Bounds	N/A
Cut Points	N/A
Zero Bin	N/A
Bin Count	2
Parent Variables	L, A
Children Variables	C_2, v_1
Total Bin Count	64
Value Equivalents	1 = 1200 code
	2 = Discrete code

TABLE A.5

Details of Aircraft Category for AC2 (C_2)

Characteristic	Variable Values
Defined Time/Period	Full Encounter Duration
Used Time/Period	Full Encounter Duration
Bounds	N/A
Cut Points	N/A
Zero Bin	N/A
Bin Count	2
Parent Variables	L, A, C_1
Children Variables	v_2
Total Bin Count	128
Value Equivalents	1 = 1200 code
	2 = Discrete code

TABLE A.6

Details of True Airspeed for AC1 (v_1)

Characteristic	**Variable Values**
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	50, 600 (knots)
Cut Points	100, 150, ..., 550 (knots)
Zero Bin	N/A
Bin Count	11
Parent Variables	L, $\dot{h}_1$, C_1, v_2
Children Variables	$\dot{v}_1$, $\dot{\psi}_1$, β, hmd
Total Bin Count	32912
Value Equivalents	N/A

TABLE A.7

Details of True Airspeed for AC2 (v_2)

Characteristic	**Variable Values**
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	50, 600 (knots)
Cut Points	100, 150, ..., 550 (knots)
Zero Bin	N/A
Bin Count	11
Parent Variables	L, $\dot{h}_2$, C_2
Children Variables	v_1, $\dot{v}_2$, $\dot{\psi}_2$, β, hmd
Total Bin Count	2992
Value Equivalents	N/A

TABLE A.8

Details of Vertical Rate for AC1 ($\dot{h}_1$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-6000, 6000 (feet/minute)
Cut Points	±5000, ±4000, ±3000, ±2500, ±2000, ±1500, ±1000, ±400 (feet/minute)
Zero Bin	[-400, 400) (feet/minute)
Bin Count	17
Parent Variables	L
Children Variables	v_1, $\dot{v}_1$, vmd
Total Bin Count	136
Value Equivalents	N/A

TABLE A.9

Details of Vertical Rate for AC2 ($\dot{h}_1$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-6000, 6000 (feet/minute)
Cut Points	±5000, ±4000, ±3000, ±2500, ±2000, ±1500, ±1000, ±400 (feet/minute)
Zero Bin	[-400, 400) (feet/minute)
Bin Count	17
Parent Variables	L
Children Variables	v_2, $\dot{v}_2$, vmd
Total Bin Count	136
Value Equivalents	N/A

TABLE A.10

Details of Turn Rate for AC1 ($\dot{\psi}_1$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-8, 8 (degrees/second)
Cut Points	±7, ±6, ±5, ±3.5, ±2, ±1, ±0.25 (degrees/second)
Zero Bin	[-0.25, 0.25) (degrees/second)
Bin Count	15
Parent Variables	v_1, $\dot{v}_1$
Children Variables	$\dot{\psi}_2$
Total Bin Count	1815
Value Equivalents	N/A

TABLE A.11

Details of Turn Rate for AC2 ($\dot{\psi}_2$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-8, 8 (degrees/second)
Cut Points	±7, ±6, ±5, ±3.5, ±2, ±1, ±0.25 (degrees/second)
Zero Bin	[-0.25, 0.25) (degrees/second)
Bin Count	15
Parent Variables	v_2, $\dot{v}_2$, $\dot{\psi}_1$
Children Variables	None
Total Bin Count	27225
Value Equivalents	N/A

TABLE A.12

Details of Acceleration for AC1 ($\dot{v}_1$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-5, 5 (knots/second)
Cut Points	±4, ±3, ±2, ±1, ±0.25 (knots/second)
Zero Bin	[0.25,0.25) (knots/second)
Bin Count	11
Parent Variables	L, $\dot{h}_1$, v_1
Children Variables	$\dot{\psi}_1$
Total Bin Count	16456
Value Equivalents	N/A

TABLE A.13

Details of Acceleration for AC2 ($\dot{v}_2$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-5, 5 (knots/second)
Cut Points	±4, ±3, ±2, ±1, ±0.25 (knots/second)
Zero Bins	[0.25,0.25) (knots/second)
Bin Count	11
Parent Variables	L, $\dot{h}_2$, v_2
Children Variables	$\dot{\psi}_2$
Total Bin Count	16456
Value Equivalents	N/A

TABLE A.14

Details of Approach Angle (β)

Characteristic	Variable Values
Defined Time/Period	TCA
Used Time/Period	TCA
Bounds	0, 360 (degrees)
Cut Points	30, 60, ..., 330 (degrees)
Zero Bins	N/A
Bin Count	12
Parent Variables	v_1, v_2
Children Variables	χ
Total Bin Count	1452
Value Equivalents	N/A

TABLE A.15

Details of Bearing (χ)

Characteristic	Variable Values
Defined Time/Period	TCA
Used Time/Period	TCA
Bounds	0, 360 (degrees)
Cut Points	90, 270 (angles 270-360/0-90 are a bin) (degrees)
Zero Bins	N/A
Bin Count	2
Parent Variables	β, hmd
Children Variables	χ
Total Bin Count	64
Value Equivalents	1 = chi $<$ 90 or chi $\geq$ 270 (in front)
	2 = 90 $\geq \chi <$ 270 (behind)

TABLE A.16

Details of Horizontal Miss Distance (hmd)

Characteristic	Variable Values
Defined Time/Period	TCA
Used Time/Period	TCA
Bounds	0, 3 (NM)
Cut Points	500 (feet)
	0.5, 1, 1.5, 2, 2.5 (NM)
Zero Bins	N/A
Bin Count	7
Parent Variables	L, υ_1, υ_2, vmd,
Children Variables	χ
Total Bin Count	243936
Value Equivalents	N/A

TABLE A.17

Details of Vertical Miss Distance (vmd)

Characteristic	Variable Values
Defined Time/Period	TCA
Used Time/Period	TCA
Bounds	0, 6000 (feet)
Cut Points	100, 200, ..., 1000, (feet)
	1200, 1400, ..., 5800 (feet)
Zero Bins	N/A
Bin Count	36
Parent Variables	L, $\dot{h}_1$, $\dot{h}_2$
Children Variables	hmd
Total Bin Count	83232
Value Equivalents	N/A

TABLE A.18

Details of Vertical Rate for AC1 at (t+1) ($\dot{h}_1(t{+}1)$)

Characteristic	Variable Values
Defined Time/Period	Time 1 - 120
Used Time/Period	Time 1 - 120
Bounds	-6000, 6000 (feet/minute)
Cut Points	±5000, ±4000, ±3000, ±2500, ±2000, ±1500, ±1000, ±400 (feet/minute)
Zero Bin	[-400, 400) (feet/minute)
Bin Count	17
Parent Variables	L, $\dot{h}_1(t)$
Children Variables	N/A
Total Bin Count	2312
Value Equivalents	N/A

TABLE A.19

Details of Vertical Rate for AC2 at (t+1) ($\dot{h}_2(t+1)$)

Characteristic	Variable Values
Defined Time/Period	Time 1 - 120
Used Time/Period	Time 1 - 120
Bounds	-6000, 6000 (feet/minute)
Cut Points	±5000, ±4000, ±3000, ±2500, ±2000, ±1500, ±1000, ±400 (feet/minute)
Zero Bin	[-400, 400) (feet/minute)
Bin Count	17
Parent Variables	L, $\dot{h}_2(t)$
Children Variables	N/A
Total Bin Count	2312
Value Equivalents	N/A

TABLE A.20

Details of Turn Rate for AC1 at (t+1)($\dot{\psi}_1(t+1)$)

Characteristic	Variable Values
Defined Time/Period	Time 1 - 120
Used Time/Period	Time 1 - 120
Bounds	-8, 8 (degrees/second)
Cut Points	±7, ±6, ±5, ±3.5, ±2, ±1, ±0.25 (degrees/second)
Zero Bin	[-0.25, 0.25) (degrees/second)
Bin Count	15
Parent Variables	L, $\dot{\psi}_1(t)$
Children Variables	N/A
Total Bin Count	30600
Value Equivalents	N/A

TABLE A.21

Details of Turn Rate for AC2 at (t+1) ($\dot{\psi}_2(t+1)$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-8, 8 (degrees/second)
Cut Points	±7, ±6, ±5, ±3.5, ±2, ±1, ±0.25 (degrees/second)
Zero Bin	[-0.25, 0.25) (degrees/second)
Bin Count	15
Parent Variables	L, $\dot{\psi}_2(t)$
Children Variables	None
Total Bin Count	30600
Value Equivalents	N/A

TABLE A.22

Details of Acceleration for AC1 at (t+1) ($\dot{v}_1(t+1)$)

Characteristic	Variable Values
Defined Time/Period	Time 0
Used Time/Period	Time 0
Bounds	-5, 5 (knots/second)
Cut Points	±4, ±3, ±2, ±1, ±0.25 (knots/second)
Zero Bin	[0.25,0.25) (knots/second)
Bin Count	11
Parent Variables	L, $\dot{h}_1(t+1)$, v_1, $\dot{v}_1(t)$, $\dot{\psi}_1(t+1)$
Children Variables	N/A
Total Bin Count	2715240
Value Equivalents	N/A

TABLE A.23

Details of Acceleration for AC2 at (t+1) ($\dot{v}_2(t+1)$)

Characteristic	Variable Values
Defined Time/Period	Time 1 - 120
Used Time/Period	Time 1 - 120
Bounds	-5, 5 (knots/second)
Cut Points	±4, ±3, ±2, ±1, ±0.25 (knots/second)
Zero Bins	[0.25,0.25) (knots/second)
Bin Count	11
Parent Variables	L, $\dot{h}_2(t+1)$, v_2, $\dot{v}_2(t)$, $\dot{\psi}_2(t+1)$
Children Variables	N/A
Total Bin Count	2715240
Value Equivalents	N/A

B. ENCOUNTER GENERATION

This section will describe in detail how an encounter is generated from the ECEM model.

B.1 INITIALIZATION NETWORK SAMPLING

To genreate an encounter the Initialization Network must be sampled from initially to produce the values needed for the Trajectory Generation Network. The Initialization Network itself must also be sampled in order.

From the ECEM the ordering for assessing variables in the Initialization Network can be achieved by taking a topological sort of the graphical structure of the network, shown in Table B.1. The graphical structure of the network is shown below in Figure B.1.

TABLE B.1

Matrix Representation of the Initialization Network Structure

0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0
0	0	0	1	0	0	1	0	0	1	0	0	0	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

One ordered sort would be:

$X_2, X_{12}, X_{11}, X_{16}, X_1, X_5, X_6, X_8, X_{10}, X_7, X_{15}, X_9, X_{13}, X_{14}, X_4, X_3$

Where each X corresponds to the variables in the order found in the parameter file. This results in the variable sampling order:

$L, \dot{h}_2, \dot{h}_1, vmd, A, C_1, C_2, v_2, \dot{v}_2, v_1, hmd, \dot{v}_1, \dot{\psi}_1, \dot{\psi}_2, \beta, \chi$

Given this order, the first variable to be sampled is L. The following equation shows how to selcted this random sample.

$$P(X_i = k|\pi_{ij}, D, G) = \frac{a_{ijk} + N_{ijk}}{\sum_{k'=1}^{r_i}(aijk' + N_{ijk'})}. \quad \text{(B.3)}$$

Where:

a_{ijk} is the objective prior (set to 1).

N_{ijk} is the frequency of this characteristic in the data.

Therefore, the probability of selecting a particular bin for L is proportion to the sum of the objective prior (1) and the bin count. Bin counts are accessible in the model parameters, the counts for L are shown in Table B.2.

The probability of selecting a particular bin from L is done using Equation (B.3).

$$\begin{aligned}P(L=1) = (20069+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.0036\end{aligned} \tag{B.4a}$$

$$\begin{aligned}P(L=2) = (460354+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.0832\end{aligned} \tag{B.4b}$$

$$\begin{aligned}P(L=3) = (680549+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.1231\end{aligned} \tag{B.4c}$$

$$\begin{aligned}P(L=4) = (1220985+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.2208\end{aligned} \tag{B.4d}$$

$$\begin{aligned}P(L=5) = (1453888+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.2629\end{aligned} \tag{B.4e}$$

$$\begin{aligned}P(L=6) = (673961+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.1219\end{aligned} \tag{B.4f}$$

$$\begin{aligned}P(L=7) = (963531+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.1742\end{aligned} \tag{B.4g}$$

$$\begin{aligned}P(L=8) = (57328+1)/(&20069+1+460354+1+680549+1+1220985+1\\&+1453888+1+673961+1+963531+1+57328+1) = 0.0104\end{aligned} \tag{B.4h}$$

A random number generator is used to choose the bin for L, for this example L=4 is chosen.

Next the second variable $\dot{\mathrm{h}}_1$ is chosen given that L=4. Bin counts for $\dot{\mathrm{h}}_1$ are shown in Table B.3, counts where L=4 are in bold. Bin probabilities are shown in Table B.4, bin probabilties where L=4 are in bold.

Each sampled variable is performed given the values of determined variables until the full Initialization Network has been sampled. After bins are selected for each variable, select variables have specific values sampled from within the sampled bins in a uniform distribution. Bins whose range span zero have zero returned as the value for that variable.

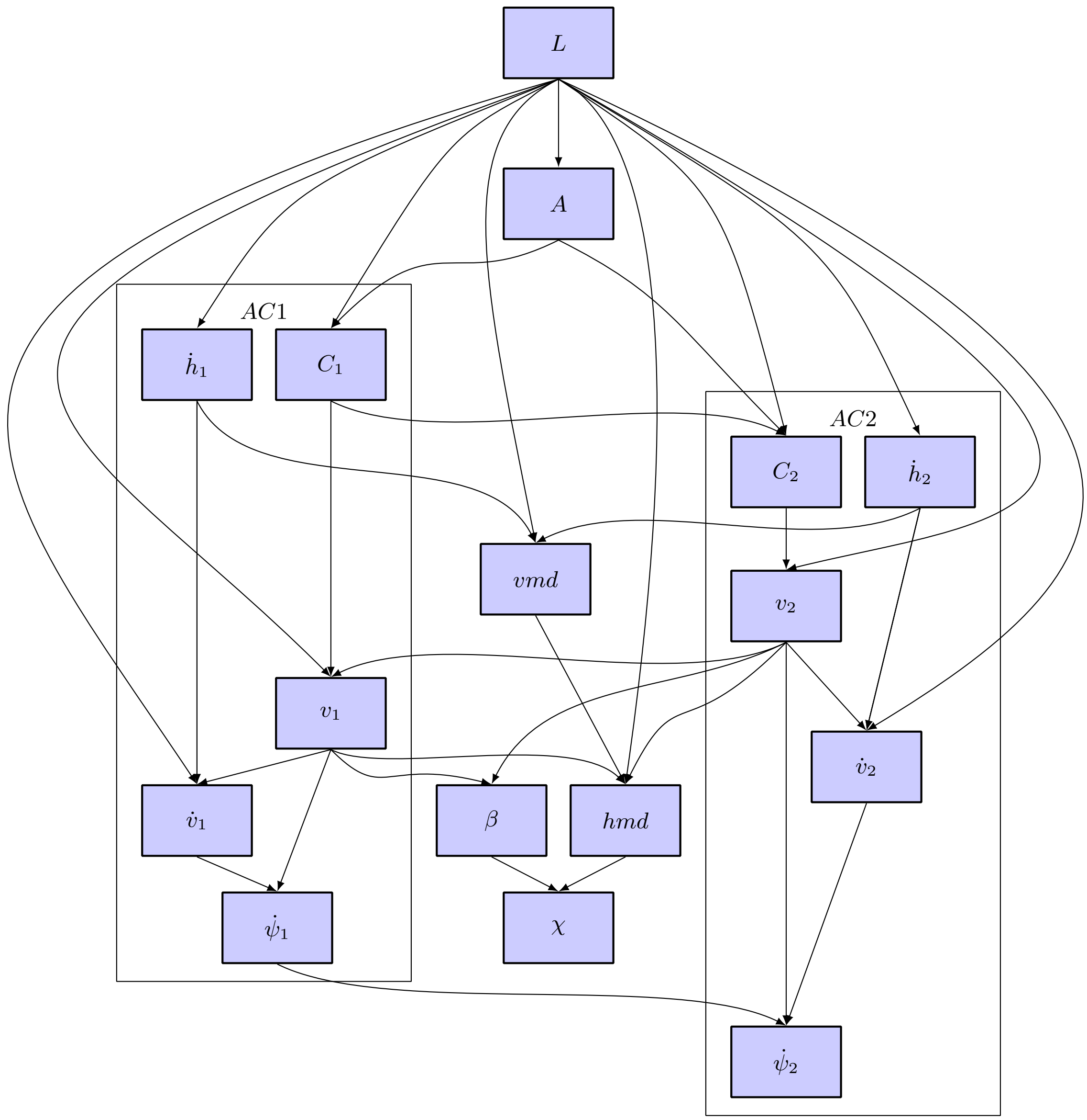

Figure B.1. The graphical representation of the the Bayesian Network for the initialization step. Rectangles represent system variables, arrows indicate dependencies between variables.

TABLE B.2

Bin Probabilities for Altitude Layer

L	**Bin Counts**
1	20069
2	460354
3	680549
4	1220985
5	1453888
6	673961
7	963531
8	57328

B.2 TRAJECTORY GENERATION NETWORK SAMPLING

Sampling from the Trajectory Generation Network is much like the Initialization Network sampling. Similarly the Trajectory Generation network must be sorted and sampled in order. The matrix that is used to represent the Trajectory Generation Network is shown in Table B.5, the graphical representation of the Trajectory Generation Network is shown in Figure B.2.

When using the Trajectory Generation Network only new variables are sampled, these leverage the bins selected by the Initialization Network sampling that occurred first. The variable values produced by the Initialization Network sampling are for Time 0, this first sampling from the Trajectory Generation Network produces values for those same variables at Time 1. For a 120 second Trajectory the Trajectory Generation Network must be sampled 119 times.

Similar to the Initialization Network sampling, the Trajectory Generation Network will first have bins sampled and second have values subsampled. However, for the Trajectory Generation Network sampling an exact value does not occur with every bin sampled. If the same bin is selected as the previous iteration the rare of resampling is dependent on the Resampling Probability. Resampling Probabilities are shown in Table B.6. Sampling within a bin for every bin selected would produce excessive variation in the trajectory. Sampling with the bins is done with a uniform distribution. Zero values will be returned if a Zero Bin is selected, and variable limits will determine the value limits for the first and last bins.

TABLE B.3

Bin Counts for Vertical Rate for AC1

$\dot{h}_1$ **Bin**	***L*=1**	***L*=2**	***L*=3**	***L*=4**	***L*=5**	***L*=6**	***L*=7**	***L*=8**
1	4	75	85	**138**	287	522	921	79
2	3	46	113	**158**	315	692	1169	85
3	6	161	364	**724**	1671	3127	3813	214
4	11	431	1136	**2452**	5354	9150	8583	415
5	30	1746	5495	**11196**	21335	28901	20355	689
6	219	6238	19932	**45527**	73786	64091	26831	857
7	1659	25057	57296	**126597**	166494	101960	40099	1489
8	8430	101792	137992	**216103**	231179	93376	52396	2502
9	9484	310956	430599	**691508**	674744	237689	655520	44151
10	195	13178	24408	**57840**	53892	22367	45265	2754
11	17	548	2214	**21159**	38137	25939	47956	1985
12	9	88	637	**19299**	47215	28342	31805	1208
13	2	23	218	**15299**	57451	25738	16220	544
14	0	7	40	**8908**	45244	17020	6762	180
15	0	6	15	**3511**	26070	11174	3679	96
16	0	2	2	**381**	6264	2224	1104	39
17	0	0	3	**185**	4450	1649	1053	41

TABLE B.4

Bin Probabilities for Vertical Rate for AC1

$\dot{h}_1$ **Bin**	***L*=1**	***L*=2**	***L*=3**	***L*=4**	***L*=5**	***L*=6**	***L*=7**	***L*=8**
1	0.0002	0.0002	0.0001	**0.0001**	0.0002	0.0008	0.0010	0.0014
2	0.0002	0.0001	0.0002	**0.0001**	0.0002	0.0010	0.0012	0.0015
3	0.0003	0.0004	0.0005	**0.0006**	0.0012	0.0046	0.0040	0.0037
4	0.0006	0.0009	0.0017	**0.0020**	0.0037	0.0136	0.0089	0.0073
5	0.0015	0.0038	0.0081	**0.0092**	0.0147	0.0429	0.0211	0.0120
6	0.0110	0.0136	0.0293	**0.0373**	0.0508	0.0951	0.0278	0.0150
7	0.0826	0.0544	0.0842	**0.1037**	0.1145	0.1513	0.0416	0.0260
8	0.4197	0.2211	0.2028	**0.1770**	0.1590	0.1385	0.0544	0.0436
9	0.4722	0.6754	0.6327	**0.5663**	0.4641	0.3527	0.6803	0.7699
10	0.0098	0.0286	0.0359	**0.0474**	0.0371	0.0332	0.0470	0.0480
11	0.0009	0.0012	0.0033	**0.0173**	0.0262	0.0385	0.0498	0.0346
12	0.0005	0.0002	0.0009	**0.0158**	0.0325	0.0421	0.0330	0.0211
13	0.0001	0.0001	0.0003	**0.0125**	0.0395	0.0382	0.0168	0.0095
14	0.0000	0.0000	0.0001	**0.0073**	0.0311	0.0253	0.0070	0.0032
15	0.0000	0.0000	0.0000	**0.0029**	0.0179	0.0166	0.0038	0.0017
16	0.0000	0.0000	0.0000	**0.0003**	0.0043	0.0033	0.0011	0.0007
17	0.0000	0.0000	0.0000	**0.0002**	0.0031	0.0024	0.0011	0.0007

TABLE B.5

Matrix Representation of the Trajectory Generation Network Structure

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

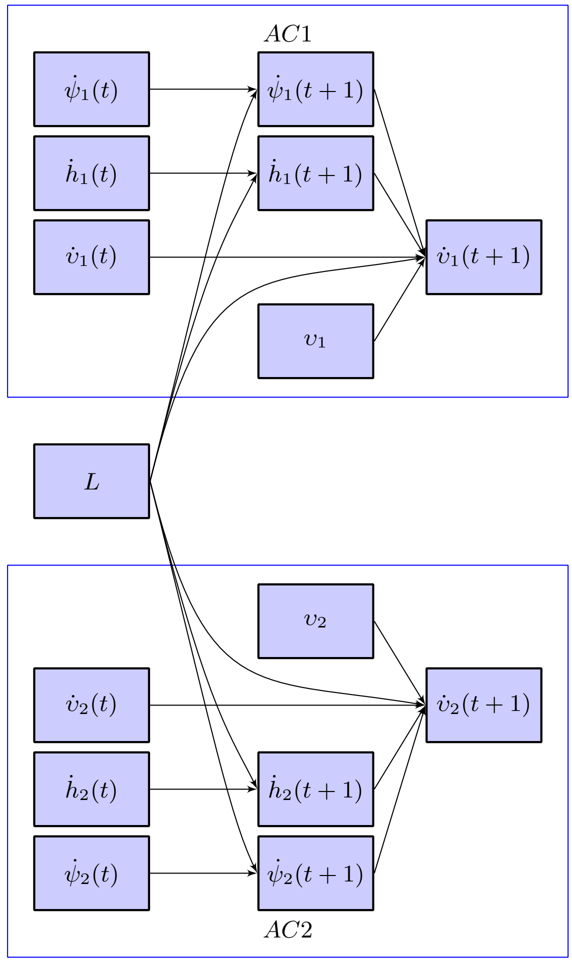

Figure B.2. The graphical representation of the the Bayesian Network for the trajectory generation step. Rectangles represent system variables, arrows indicate dependencies between variables.

TABLE B.6

Resampling Rates for Trajectory Generation Dynamic Variables

Variable	Resampling Probability
Vertical Rate AC1 ($\dot{h}_1$):	0.0820
Vertical Rate AC2 ($\dot{h}_2$):	0.0655
Turn Rate AC1 ($\dot{\psi}_1$):	0.0898
Turn Rate AC2 ($\dot{\psi}_2$):	0.1035
Acceleration AC1 ($\dot{v}_1$):	0.0248
Acceleration AC2 ($\dot{v}_2$):	0.0227

This page intentionally left blank.

C. SAMPLE ENCOUNTERS

This section details sample encounters that may be used for validation. These samples may be requested from MIT LL.

Encounters in these files are 120 seconds in duration, and were generated by the sampling processes and the ECEM described earlier in this report. TCA for these encounters occur at 110 seconds into the encounter.

Values for the Initialization Network and values for the Trajectory Generation Network are stored in separate files. The format of the encounters in the Initialization network is to have one encounter per row. Each row contains the following variable values in order:

1. Encounter ID
2. Airspace Class A
3. Altitude Layer L
4. Bearing χ
5. Approach Angle β
6. Aircraft Category AC1 C_1
7. Aircraft Category AC2 C_2
8. True Airspeed AC1 v_1
9. True Airspeed AC2 v_2
10. Acceleration AC1 $\dot{v}_1$
11. Acceleration AC2 $\dot{v}_2$
12. Vertical Rate AC1 $\dot{h}_1$
13. Vertical Rate AC2 $\dot{h}_2$
14. Turn Rate AC1 $\dot{\psi}_1$
15. Turn Rate AC2 $\dot{\psi}_2$
16. Horizontal Miss Distance hmd
17. Vertical Miss Distance vmd

The Trajectory Generation Network file has a row for each encounter at each time. The format for a row in these files is:

1. Encounter ID
2. Time (seconds into encounter)
3. Vertical Rate AC1 $\dot{h}_1$
4. Vertical Rate AC2 $\dot{h}_2$
5. Turn Rate AC1 $\dot{\psi}_1$
6. Turn Rate AC2 $\dot{\psi}_2$
7. Acceleration AC1 $\dot{v}_1$
8. Acceleration AC2 $\dot{v}_2$

To completely model an encounter a single row is needed from the Initialization file, and the 120 rows pertaining to the encounter must be collected from the Trajectory Generation file. Additionally, the initial values for vertical rates, turn rates, and acceleration will be the same in the Trajectory Generation file for time 0 as the Initialization Network values.

D. ENCOUNTER IDENTIFICATION CRITERIA (ENCOUNTER FILTER)

This appendix explains how candidate encounters from surveillance data are chosen for inclusion when building the model. The process differs from the prior version of the correlated encounter model, due to the different intent and scope of this version.

D.1 ASSUMPTIONS

For this model, an encounter is defined as a situation involving two aircraft whose trajectories would likely trigger a collision avoidance or DAA system. As explained below, the definition is intended to be more inclusive or lenient than in the prior version of the correlated encounter model [6, 8].

Encounters are identified using a modified version of the alerting logic used in RTCA SC-228 modeling and simulation work [1]. The alerting logic captures encounters of interest when designing a system to meet SC-228 requirements. However, this model is intended to serve a wider scope of modeling needs and wider community of users, so the parameters are adjusted accordingly.

Encounter identification is intended to be relatively inclusive and err on the side of including more, rather than fewer encounters, which is accomplished by choosing larger & more lenient threshold parameter values. There are several reasons for this choice:

1. Unknown future changes to DAA system requirements (or other developments) may require a wider modeling scope
2. Less effort will be needed in the future to reduce the model's scope than to expand it, hence it is better to begin with a wider scope
3. Errors in surveillance data require an extra buffer to ensure encounters of interest are captured

D.2 ENCOUNTER IDENTIFICATION CRITERIA

The encounter identification condition takes the form of the SC-228 Non-Hazard Zone (NHZ) boundary definition, but with expanded, more lenient, threshold parameters. It is intended to identify situations passing into a expanded version of the NHZ boundary, which means it captures those passing into the Hazard zone or May-Alert zone, as defined by SC-228.

The condition in equation D.5 defines the criteria for encounters to be included in the model:

$$\tau_{mod} \leq \tau^*_{mod} \underline{and} HMD_p \leq HMD^* \underline{and} h_r \leq V \tag{D.5}$$

Where $V = max(V_{mod}, V_{mod} - \dot{h}_r \tau^*_{mod})$ is the vertical threshold boundary for the NHZ, and V_{mod} defines the minimum value of V. Note that $V > V_{mod}$ if aircraft are converging vertically, which makes the test more inclusive in that case–helping to capture pairs that appear likely to cross in altitude.

The following variables are used:

h_r = vertical separation

HMD_p = projected Horizontal Miss Distance (HMD), at linearly-projected TCA.

$\dot{h}_r$ = vertical closure rate (negative if converging)

τ_{mod} = modified Tau value as defined by SC-228.

The following parameters values are used:

τ^*_{mod} = 165 seconds

HMD^* = 3.0 NM

V_{mod} = 1200 feet

These values are chosen as 50% greater than those from the original SC-228 NHZ definition, to ensure encounters of interest are captured for model building. The τ^*_{mod} parameter is consistent with the 120 second encounter duration used in modeling.

The condition is tested at each timestep of a candidate encounter, and if it is satisfied at one or more timesteps, the encounter is included in the model.

This processing uses archived data (not real-time), therefore the actual HMD is known. However, the test uses linearly-projected HMD to mimic real-time alerting logic behavior.

Note that this condition captures some aircraft pairs typically considered safely separated: for example, level flight under 1000 feet vertical separation. As discussed in earlier sections, this is desirable so that the encounter model can be used to evaluate a DAA system's behavior when aircraft are safely separated.

D.3 VALIDATION

Encounter identification criteria were tested on small subsets of surveillance data before being applied to the full set used to build the model. During model development, the process was validated by inspection and analysis of the resulting encounter sets, as described earlier in this report.

E. TRACKING AND FUSION

Converting raw radar reports into tracks that are usable for our model development is a two-stage process. The first stage involves forming local tracks from the reports associated with each sensor. The second stage involves fusing the local tracks from multiple sensors to form global tracks. This appendix provides a brief overview of tracking and fusion.

The tracking algorithms from the Mode S and ASR-9 systems [11], the two most modern sensors in the Air Traffic Control System are used. The beacon correlation algorithms come from Mode S and the primary radar correlation algorithms come from ASR-9 Processor Augmentation Card (9-PAC). Both systems are integrated to provide a consistent track, although for the purpose of this project we ignore primary only reports.

After the reports of each sensor have been correlated into local tracks, the local tracks can be fused to provide a global picture of the airspace. Fusion performs the following functions:

1. Merge tracks from multiple sensors that correspond to the same aircraft into a single global track.
2. Compute the speed and heading of each global track to permit trajectory predictions.
3. Correct sensor tracking errors that would have led to split global tracks and thus false encounters.

A *track-to-track* fusion method is used, meaning that each sensor's individual reports are tracked and then all tracks are merged [12]. The main advantages of *track-to-track* fusion over *report-to-track* fusion, in which all reports are correlated directly to global tracks, are:

1. Bias independence for velocity determination and maneuver detection.
2. Removal of short update interval velocity anomalies.
3. Reduced likelihood of forming clutter tracks.
4. Reduced likelihood of introducing incorrect data points into the track due to correlation errors.

Track merging employs position, velocity, Mode A code, and altitude as matching attributes over the entire track. Track merging for tracks with discrete codes, which are unique within an area, employs large correlation boxes for each of the matching attributes. Other tracks (1200 code or radar-only) must pass more stringent position tests and velocity tests in order to be merged together.

This fusion method works forward in time. Thus, there is often doubt in whether or not tracks from multiple radars are indeed the same track with just a few data points. In cases of doubt, tentative matches are remembered that can be upgraded to a merge after more scans of information are obtained. Merges are checked each scan and can be undone if later found to be unsatisfactory. Aircraft code changes are also accommodated, although they must be verified by other sensors in the merge set of the global track before being accepted. If only one sensor reports a code change, it is assumed that the sensor had a track swap, and the merge situation is altered accordingly.

The remainder of this section explains how local sensor tracks are added to global tracks, how to estimate track velocity as part of the fusion process, and how to filter for encounters.

E.1 ADDING LOCAL SENSOR TRACKS TO GLOBAL TRACKS

In order to facilitate fusing of tracks from multiple sensors into global tracks, the continental United States is broken into 20 NM by 20 NM bins. Every track is associated to a geographic bin. Whenever the fusion process receives a new local sensor track, or a later report for an as yet unfused local track, an attempt is made to fuse this track to an existing global track. This process is performed by comparing the new track to all neighboring global tracks. The neighboring global tracks for a local track include all global tracks in the same geographical bin as the local track plus global tracks in the surrounding bins (9 bins total). Several tests must be passed for the successful fusion of the single track to a global track:

1. The global track must not already be fused to another local track from the new track's sensor.
2. The tracks must agree on Mode A code (primary-only tracks automatically pass).
3. If the code agreement was on a discrete code, a very coarse horizontal positional test must be passed.
4. If the code agreement was on 1200 code, or no codes, a tighter positional test must be passed, as well as altitude and velocity tests. However, if only the velocity test fails, a potential fusion is declared; three successive potential fusions with the same global track results in a successful fusion.

If more than one possible global track satisfies the fusion tests, the one with the highest matching score is chosen. Existing fusion matches are checked each time a new report is received. If the tests fail for three scans in a row, the fusion is ended, and a new global track is sought for the local sensor track.

E.1.1 Code Matching Test

Normally, the code of the new track is the same as the code of the global track. However, code mismatching can complicate the fusion process. Code declaration errors due to data corruption, missing codes, and code changes due to controller action are all common. For this reason, associated with each global track is an established code and an alternate code, which is the code of the most recent report. Usually these two codes are the same. When an alternate code is different from the established code for three successive reports, however, the established code is updated to the alternative code. Reports with no beacon code are ignored in this process. However, if a track has never had been associated with a beacon code report, we consider this track to be a primary-only and give the track an established code of 0.

If both the local track the global track have a beacon code, then a successful code match is declared if any of the following statements are true:

1. The established codes match.
2. The alternate codes match.
3. One of the established codes and the other alternate code match.

However, failure is declared if the match is on code 0, and both tracks have a beacon code in the other code slot that do not match. Lastly, if one track is radar only, and the other track has a beacon code, failure is declared.

To handle local track code changes due to a track swap in the single sensor tracker, confirmation of the code change by the global track is required. If at the time of the local track code change, the global track has had an update by a different sensor's track with the old beacon code, a track swap is declared, and the local track is removed from fusion with the global track. The local track then undergoes a new fusion process.

E.1.2 Horizontal Position Matching Test

Horizontal positional matching requires agreement between the global track's most recent horizontal position $x_{\mathrm{g}}, y_{\mathrm{g}}$ and the new local track's horizontal position $x_{\mathrm{l}}, y_{\mathrm{l}}$ projected back to the time of the global track. This test is simple if both track's reports contained altitude. If a radar report contains altitude h, range ρ, azimuth θ, then the track's horizontal position x, y can be empirically determined.

If, however, the altitude of a track is unknown, then the tracks' altitude has to be assumed (or guessed) in order to derive the track's horizontal position. Simply guessing an altitude may produce a erroneous x, y position. In order to use a reasonable altitude value, the following algorithm is employed:

1. If only one track has known altitude, then convert the other track's stored ρ, θ position to x, y using the first track's altitude.
2. If neither track has known altitude (which is always true for a primary-only match), all altitudes are considered from 0 NM to 7 NM at 1 NM steps. Next, use the altitude that produces the closest positional match between the two tracks. While a smaller step size may produce more accurate estimates, it was found that 1 NM is sufficient for fusing two tracks.

Horizontal positional agreement is declared if the horizontal distance between the two tracks is less than an acceptable value:

$$\sqrt{(x_{\mathrm{g}} - x_{\mathrm{l}})^2 + (y_{\mathrm{g}} - y_{\mathrm{l}})^2} \leq \Delta r_{\max} + 3\sigma_{\mathrm{az}}\rho_g + 3\sigma_{\mathrm{az}}\rho_l \,, \tag{E.6}$$

where $\Delta r_{\max} = 20\,\mathrm{NM}$ is used for a discrete code match and $\Delta r_{\max} = 1\,\mathrm{NM}$ is used for a 1200 code or radar match. The standard deviation for horizontal position error terms σ_{az} account for positional errors due to azimuth noise in radar measurements (which is the dominant source of horizontal position error). The standard deviation of the azimuth noise is modeled as 3 milliradians for the data format of the radar feed.

E.1.3 Altitude Matching Test

Altitude matching requires agreement between the local and global track altitudes when both are known. Two comparisons are tested; the success of either test results in a match. The comparison test is:

$$\begin{aligned}
\Delta t_{\mathrm{l}} &= t_{\mathrm{g}} - t_{\mathrm{l}} \\
\Delta h_{\mathrm{l}} &= |h_{\mathrm{g}} - h_{\mathrm{l}}| \\
\Delta h_{\mathrm{l}} &\leq \Delta h_{\max} \\
\frac{\Delta h_{\mathrm{l}}}{\Delta t_{\mathrm{l}}} &\leq \Delta \dot{h}_{\max}
\end{aligned}$$

where $\Delta h_{\max} = 600\,\mathrm{ft}$ and $\Delta \dot{h}_{\max} = 100\,\mathrm{ft/s}$ is used. Since the altitude of a track can significantly change between sequential reports, both the most recent and previous local track altitudes with the most recent global track altitude update are tested. Only one of the local track altitudes is required to pass the test.

E.1.4 Velocity Matching Test

Velocity matching requires agreement between the two track headings ψ and speeds s according to the following tests:

$$\begin{aligned}
|\psi_{\mathrm{g}} - \psi_{\mathrm{l}}| &\leq \Delta \psi_{\max} \\
|s_{\mathrm{g}} - s_{\mathrm{l}}| &\leq \Delta s_{\max} \\
\frac{1}{2} \leq \; & \tfrac{s_{\mathrm{g}}}{s_{\mathrm{l}}} \leq 2
\end{aligned}$$

where $\Delta \psi_{\max} = 45°$ and $\Delta s_{\max} = 100\,\mathrm{kt}$ is used. The last test is needed for slow aircraft and clutter tracks, to prevent, for example, speeds of 20 and 110 kt from agreeing.

E.2 DETERMINING TRACK AIRSPEED AND HEADING

Determining a global track's airspeed and heading is a two step process. First, the individual sensor tracks are smoothed. Second, the individual track are averaged using relative weights that account for sensor update times and the quality of each sensor's measurement. Both alpha smoothing and curve fitting are applied to determine airspeed and heading, depending upon the track situation. Various maneuver detection algorithms are also applied, and tracking is dependent upon the current turn rate and acceleration states of the track.

E.2.1 Local Track Smoothing

First, it is required that the track has moved a minimum distance for it to be considered. If the track never moves more than 1 NM, then the track is thrown out. After the movement test is satisfied, the track's airspeed and heading are calculated from the new and previous positions. Next the local track's airspeed and heading estimates are updated using alpha smoothing.

First, the current heading estimate $\psi^{(j)}$, and its difference from the previous estimate $\psi^{(j-1)}$, are given by

$$\begin{aligned} \psi^{(j)} &= \operatorname{atan2}\left((x^{(j)} - x^{(j-1)}), (y^{(j)} - y^{(j-1)})\right) \\ \Delta\psi^{(j)} &= \psi^{(j)} - \psi^{(j-1)} \end{aligned}$$

Next, is to determine the current turn rate state S_ψ of the track:

$$S_\psi^{(n)} = \begin{cases} 2 & \text{if } \Delta\psi^{(j)} > \sigma_{\text{heading}} \\ 1 & \text{if } \Delta\psi^{(j)} > \Delta\psi_{\min} \\ -2 & \text{if } \Delta\psi^{(j)} < -\sigma_{\text{heading}} \\ -1 & \text{if } \Delta\psi^{(j)} < -\Delta\psi_{\min} \\ 0 & \text{otherwise} \end{cases} \tag{E.7}$$

where $\Delta\psi_{\min} = 3°$ and σ_{heading} is the standard deviation of the heading noise, which is calculated from the standard deviations for range and azimuth noise of the sensors. Note that a positive $\Delta\psi$ value corresponds to a right turn, while a negative value corresponds to a left turn. Next, $S_\psi^{(j)}$ is used to determine the smoothing value alpha α in Table E.1 of the individual tracks that will be used to calculate the heading of the global track at the current time.

The new track heading is finally given by:

$$\psi^{(j)} = \psi^{(j-1)} + \alpha \times \Delta\psi^{(j)}$$

and we iterate through this process for the entire track.

The process to estimate airspeed s is similar, with one important difference. If successive positions are simply connected, then the airspeed estimates will always be too high, since the aircraft will appear to "zig-zag" along the track. Thus, only the projection of the velocity vector onto the track's heading vector is used to determine the track's airspeed:

$$\begin{aligned} s^{(j)} &= \cos\left(\frac{\Delta\psi^{(j)}}{2}\right) \times \sqrt{\frac{(x^{(j)} - x^{(j-1)})^2 + (y^{(j)} - y^{(j-1)})^2}{t^{(j)} - t^{(j-1)}}} \\ \Delta s^{(j)} &= s^{(j)} - s^{(j-1)} \end{aligned}$$

TABLE E.1

Smoothing Values Depending on the Current and Previous Turn States

	Current Turn State				
Previous State	Large Left Turn (-2)	Small Left Turn (-1)	No Turn (0)	Small Right Turn (+1)	Large Right Turn (+2)
Large Left (-2)	0.7	0.7	0.4	0.5	0.5
Small Left (-1)	0.7	0.4	0.4	0.5	0.5
No Turn (0)	0.4	0.4	0.3	0.4	0.4
Small Right (+1)	0.5	0.5	0.4	0.4	0.7
Large Right (+2)	0.5	0.5	0.4	0.7	0.7

The current airspeed acceleration state S_s of the track is determined using a similar technique as we did for turn rate.

$$S_s^{(j)} = \begin{cases} 2 & \text{if } \Delta s^{(j)} > \sigma_{\text{speed}} \\ 1 & \text{if } \Delta s^{(j)} > \Delta s_{\min} \\ -2 & \text{if } \Delta s^{(j)} < -\sigma_{\text{heading}} \\ -1 & \text{if } \Delta s^{(j)} < -\Delta s_{\min} \\ 0 & \text{otherwise} \end{cases}$$

where $\Delta s_{\min} = 18\,\text{kt}$ and σ_{speed} is the standard deviation of airspeed error due to noise in range and azimuth measurements from the radar sensors. The speed smoothing and the speed alpha table rules are the same as for the heading case.

E.2.2 Global Track Smoothing

In order to determine a global track's airspeed and heading at each measurement, a weighted least squares estimation approach is used. This section describes in detail the approach for determining the tracks heading.

First, each sensor's heading estimate is assigned a weight w_i at the current time $t^{(c)}$ as follows:

$$w_i^{(c)} = \sigma_{\text{heading}}^{-1} \times \frac{t_{\max} - \frac{t_i^{(j)} + t_i^{(j-1)}}{2} - t^{(c)}}{t_{\max}}$$

where σ_{heading} is the standard deviation of the heading noise and $t_{\max} = 18\,\text{s}$ is a discounting factor that takes into account the time difference between the measurement from the sensor being considered and the time for when heading is being determined. The time $t_i^{(j)}$ corresponds to the time of the next closest measurement for sensor i with respect to the current time that the track's heading is being determined.

Next, the total turn state score for the track is determined

$$\left| \sum_{i=1}^{N} S_i \right| ,$$

where N is the number of sensors supporting the track and S_i is the current turn state value for the ith sensor defined in Equation E.7. If the turn rate score is less than N, then the track is considered to be non-turning and the current global heading is simply the weighted average of the N sensor heading estimates:

$$ {}^{(c)}_{\text{global}} = \frac{\sum_{i=1}^{N} {}^{(j)}_{i} \times w_i^{(j)}}{\sum_{i=1}^{N} w_i^{(c)}} $$

Otherwise, if the turn rate score is greater than or equal to N, then the track is considered to be in a turn. In this case, weighted least squares is used to determine a first-order relationship between time and heading.

The global track speed calculation is identical in form to the global track heading calculation.

E.3 COOPERATIVE ENCOUNTER IDENTIFICATION

Encounter identification is facilitated by having all global tracks sorted into $20 \times 20\,\text{NM}$ geographical bins. In addition, determining which sensor to employ for an encounter is guided by an ordered list of the sensor priorities associated with each bin; terminal sensor are preferred to enroute ones due to the more frequent update rate, and closer sensors are preferred to more distant ones because report accuracy is a function of range.

Two criteria are used to declare an encounter in progress:

1. **Proximity Condition:** Both tracks are currently within a sufficiently small horizontal and vertical box relative to each other.
2. **Closure Condition:** Both tracks, when projected to the point of closest horizontal approach, reach positions within a similar box, and the time to closest horizontal approach is within a sufficiently short time.

Encounter detection is performed for each global track update, at which time the track is compared to all other tracks in the appropriate bins. Proximity testing is done by predicting the bin track to the time of the updated track, using the track velocity. Closure testing predicts each track to the time of closest approach. The equation for the time of closest approach given the current positions and velocities of two tracks is

$$t_{\text{closest}} = -\frac{(x_1 - x_2)(\dot{x}_1 - \dot{x}_2) + (y_1 - y_2)(\dot{y}_1 - \dot{y}_2)}{(\dot{x}_1 - \dot{x}_2)^2 + (\dot{y}_1 - \dot{y}_2)^2} .$$

In order to initiate an encounter, several additional tests must be satisfied:

1. The sensor of the current update must be the highest priority sensor having surveillance on both tracks; otherwise encounter initiation awaits the report from that sensor.
2. Both tracks must be at least 3 scans old to prevent spurious or split tracks from being used; because the proximity filter is quite large, this delay will not prevent the finding of true encounters.
3. The bin track must have been updated within the last half scan; this rule prevents a track number change from creating an encounter, as both reports initiating the encounter must be on the same scan.
4. The tracks must be at a minimum altitude above ground to filter out ground traffic or aircraft on parallel approach paths.

Once an encounter is initiated, only the sensor forming the encounter can be used to maintain the encounter, thereby preventing inter-sensor biases from contaminating the track trajectories.

F. VISUAL ASSESSMENT GRAPHS

This section contains a selection of graphs used during the Visual Assesesment. A few graphs without any data points are excluded including:

- The breakdown of υ_1 by $\dot{h_1}$ and υ_2 where L=7 and C_1=1200 code.
- The breakdown of υ_1 by $\dot{h_1}$ and υ_2 where L=8 and C_1=1200 code.

The ordering of the included graphs is:

1. $\dot{h_1}$
2. $\dot{h_2}$
3. β
4. C_1
5. C_2
6. υ_1
7. $\dot{\upsilon}_2$

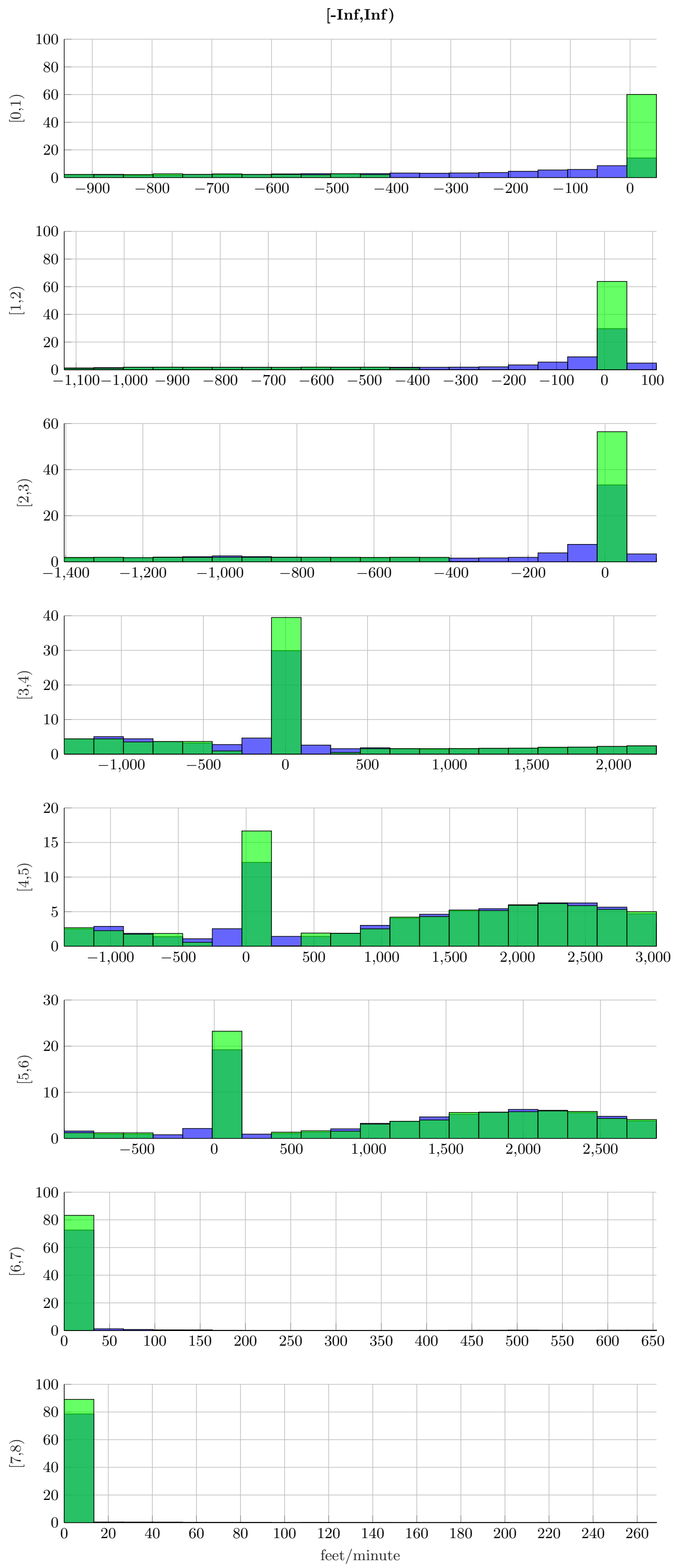

Figure F.1. The breakdown of $\dot{h}_1$ by L.

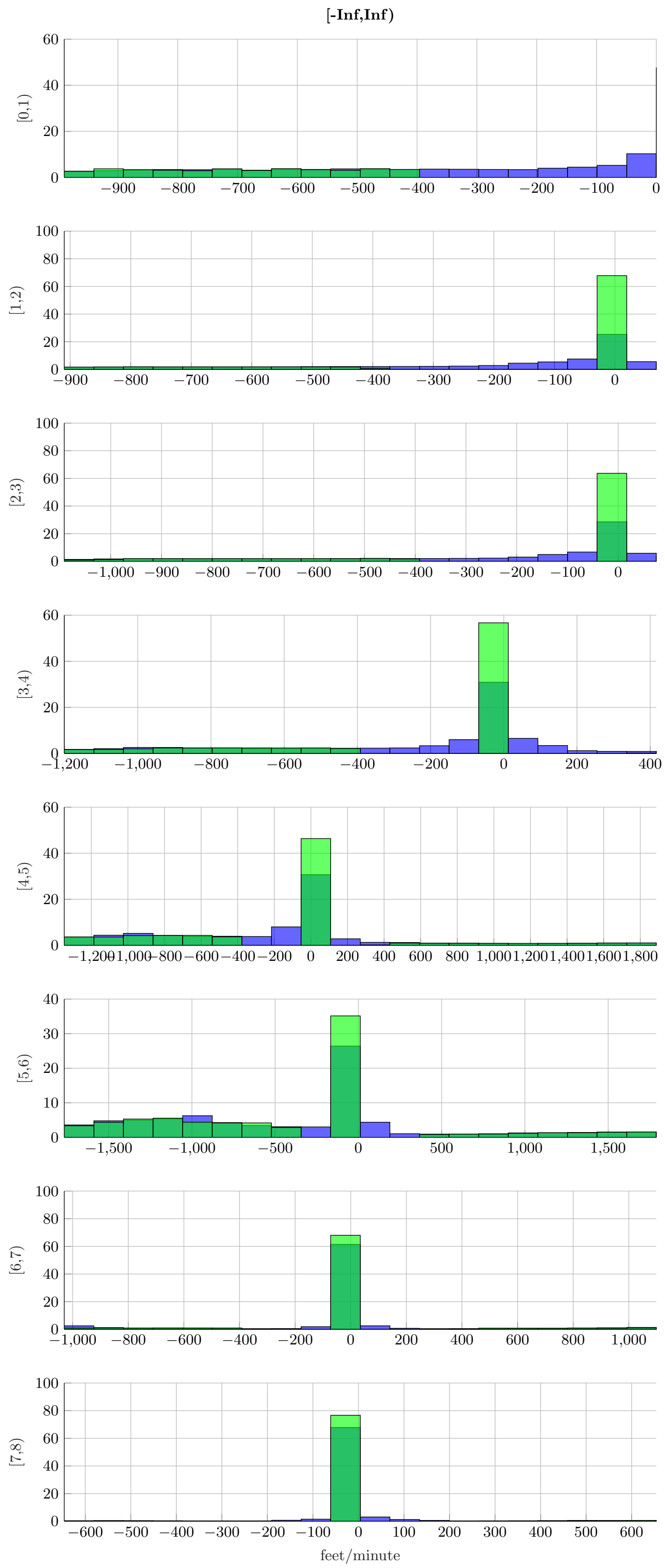

Figure F.2. The breakdown of $\dot{h}_2$ by L.

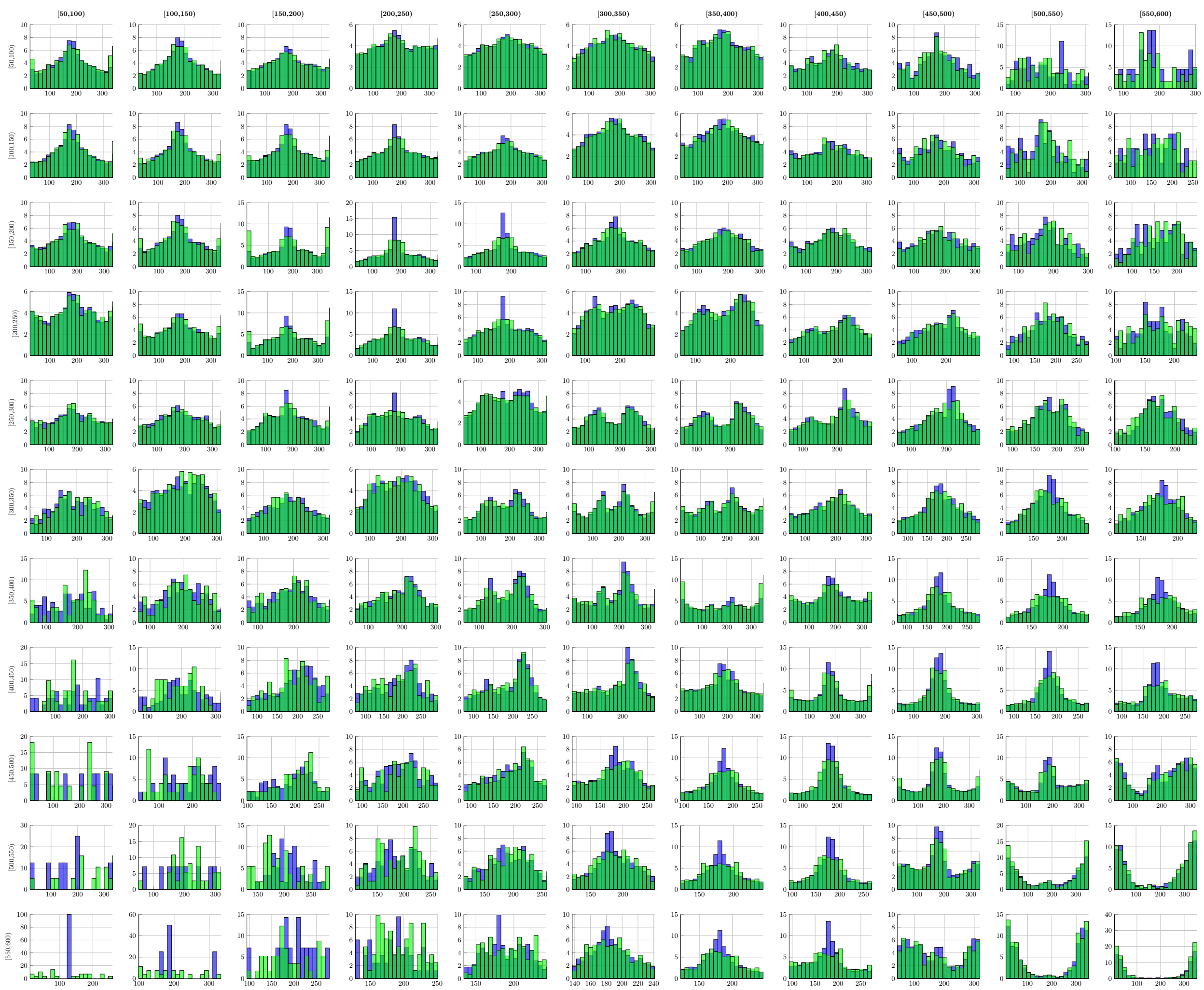

Figure F.3. The breakdown of β by υ_1 and υ_2.

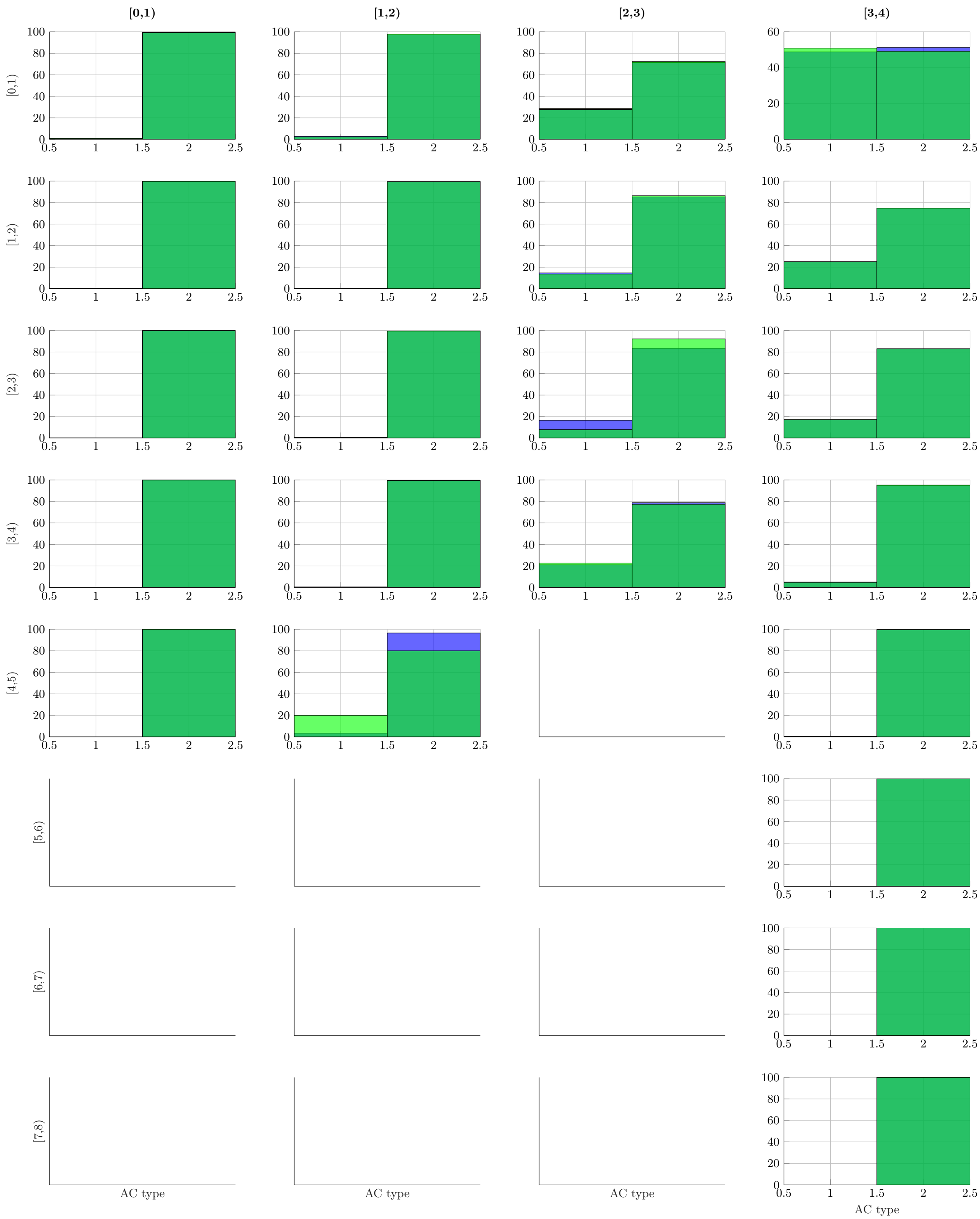

Figure F.4. The breakdown of C_1 by A and L.

Figure F.5. The breakdown of C_2 by L where C1=1200 code.

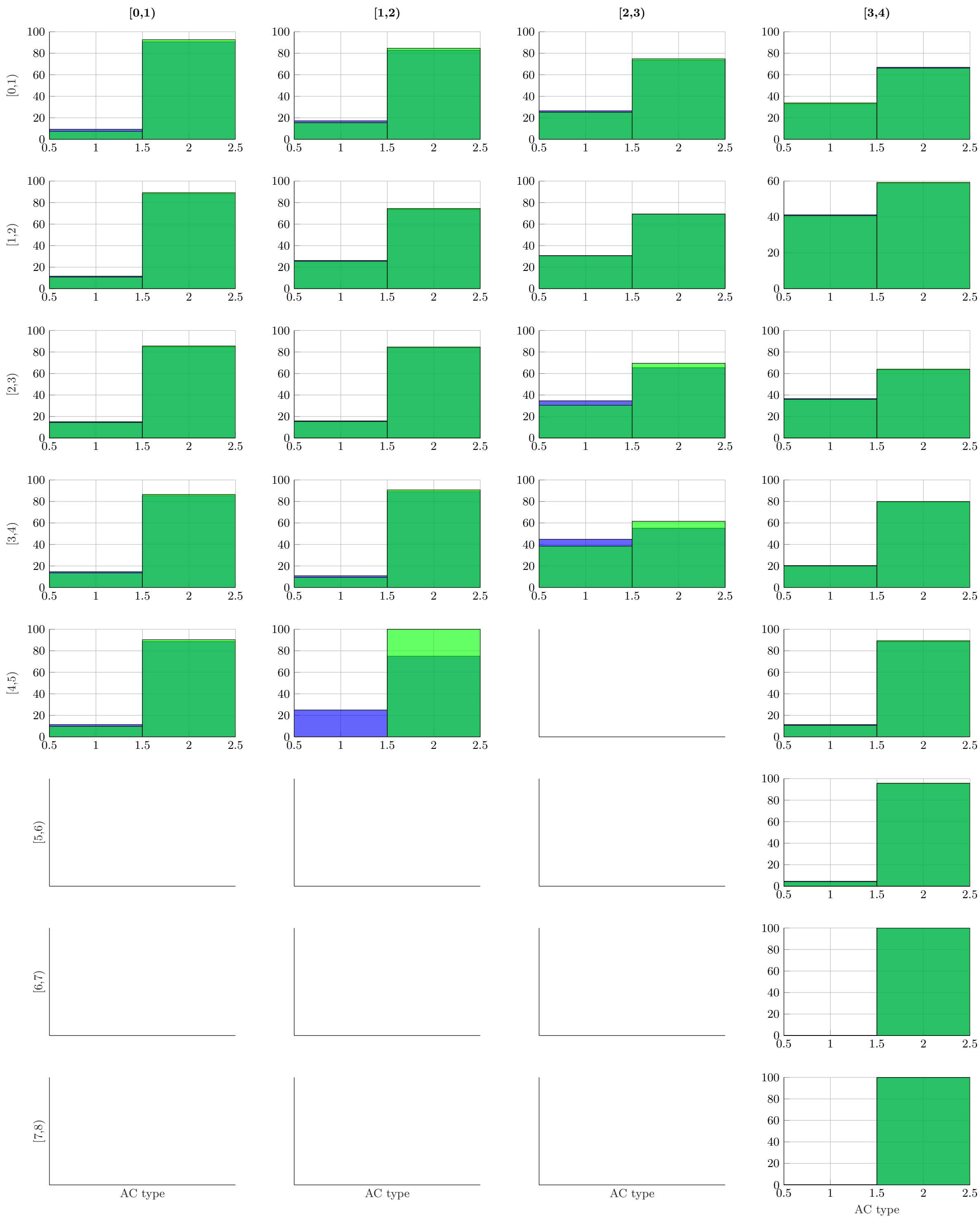

Figure F.6. The breakdown of C_2 by L where C1= discrete code.

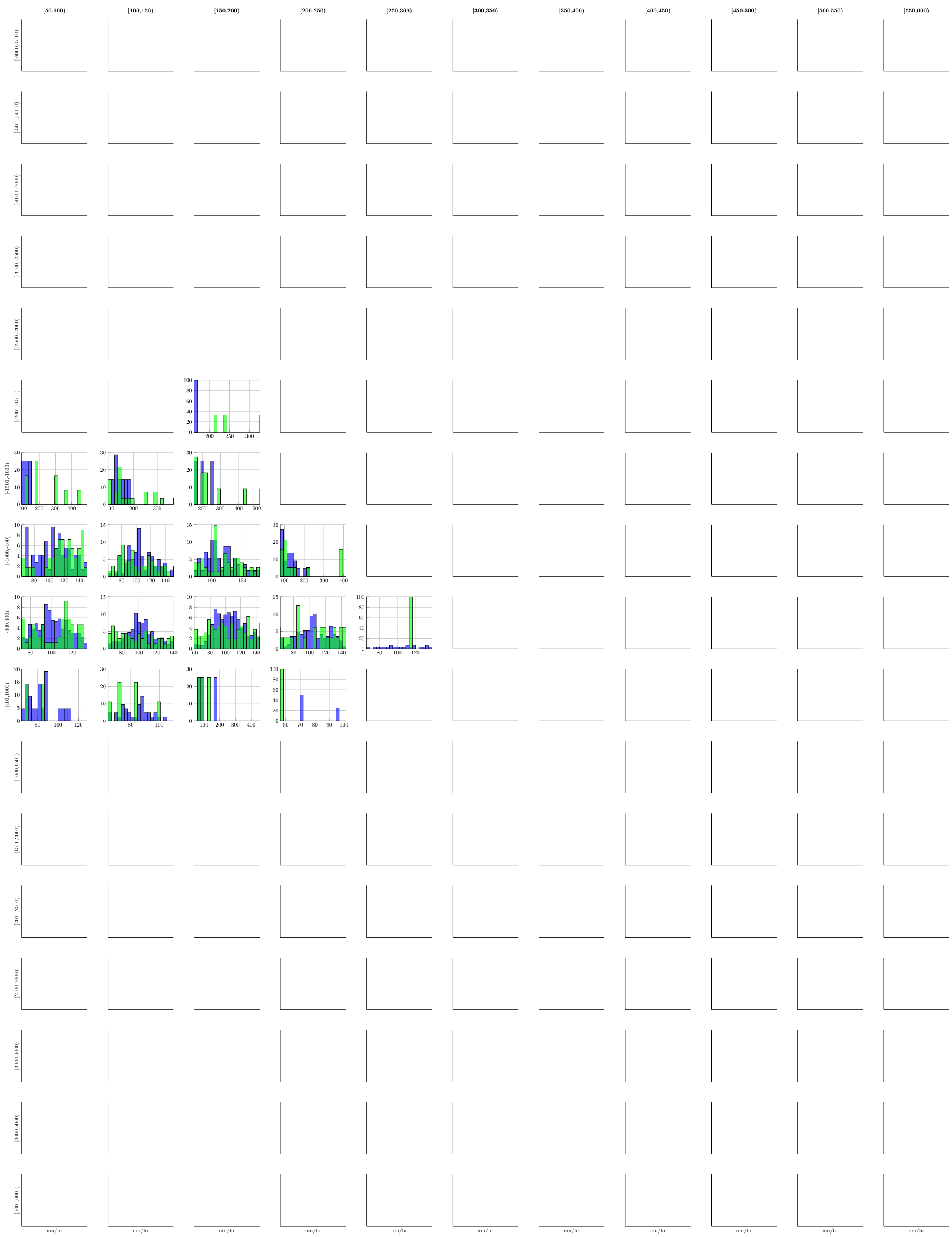

Figure F.7. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=1, C_1=1200 code.

Figure F.8. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=2, C_1=1200 code.

Figure F.9. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=4, C_1=1200 code.

Figure F.10. The breakdown of υ_1 by $\dot{h_1}$ and υ_2 where L=5, C_1=1200 code.

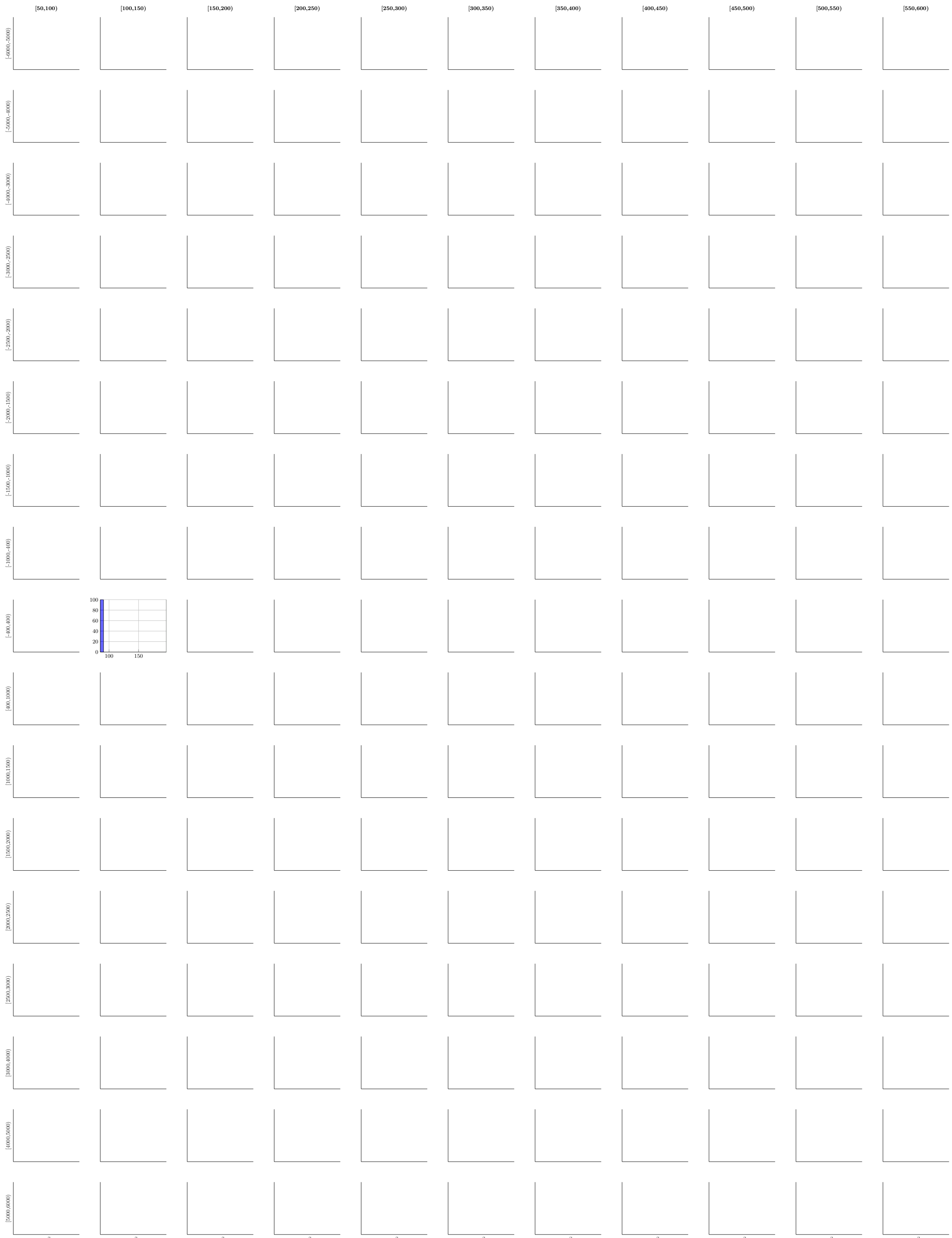

Figure F.11. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=6, C_1=1200 code.

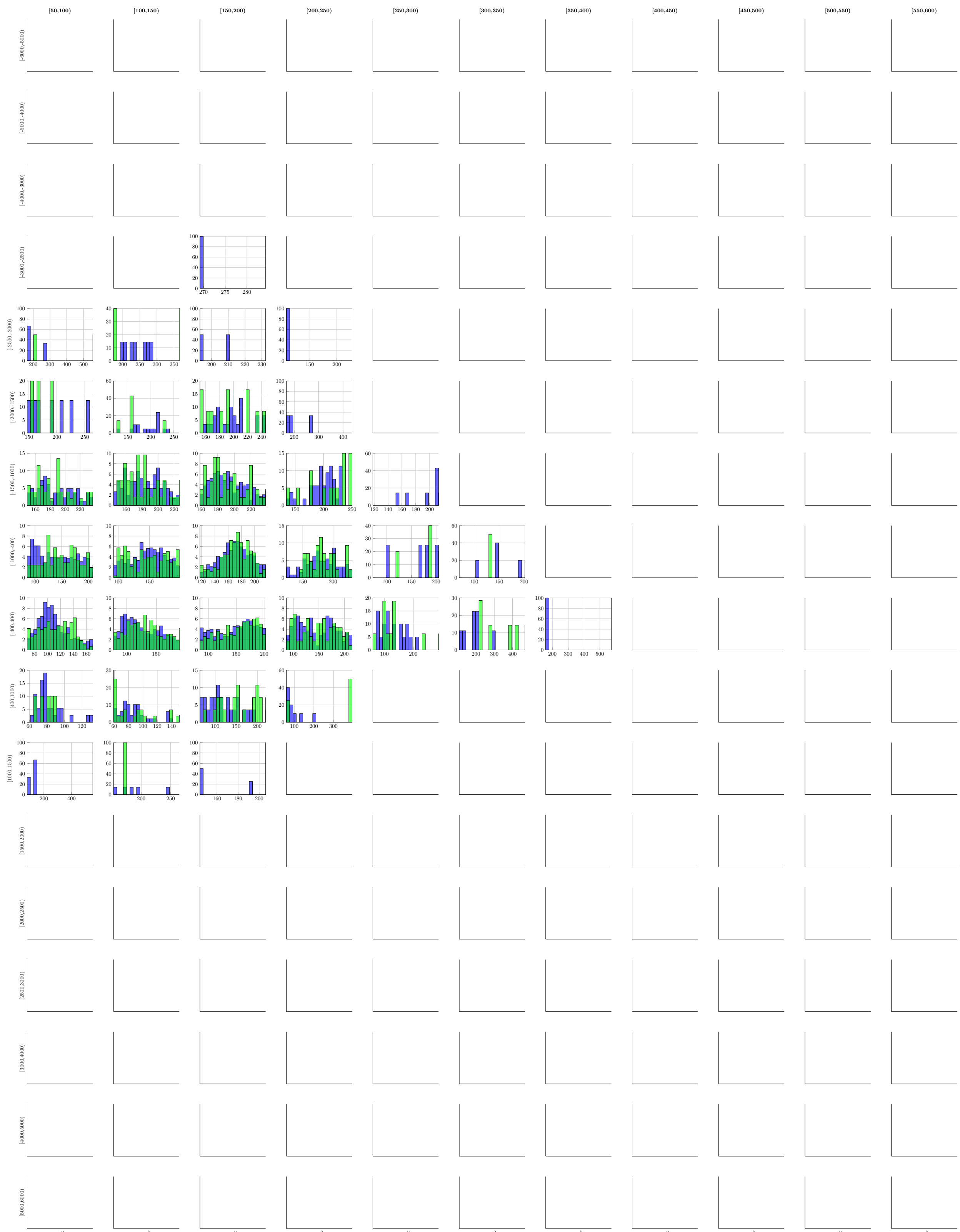

Figure F.12. The breakdown of υ_1 by $\dot{h_1}$ and υ_2 where L=1, C_1=Discrete code.

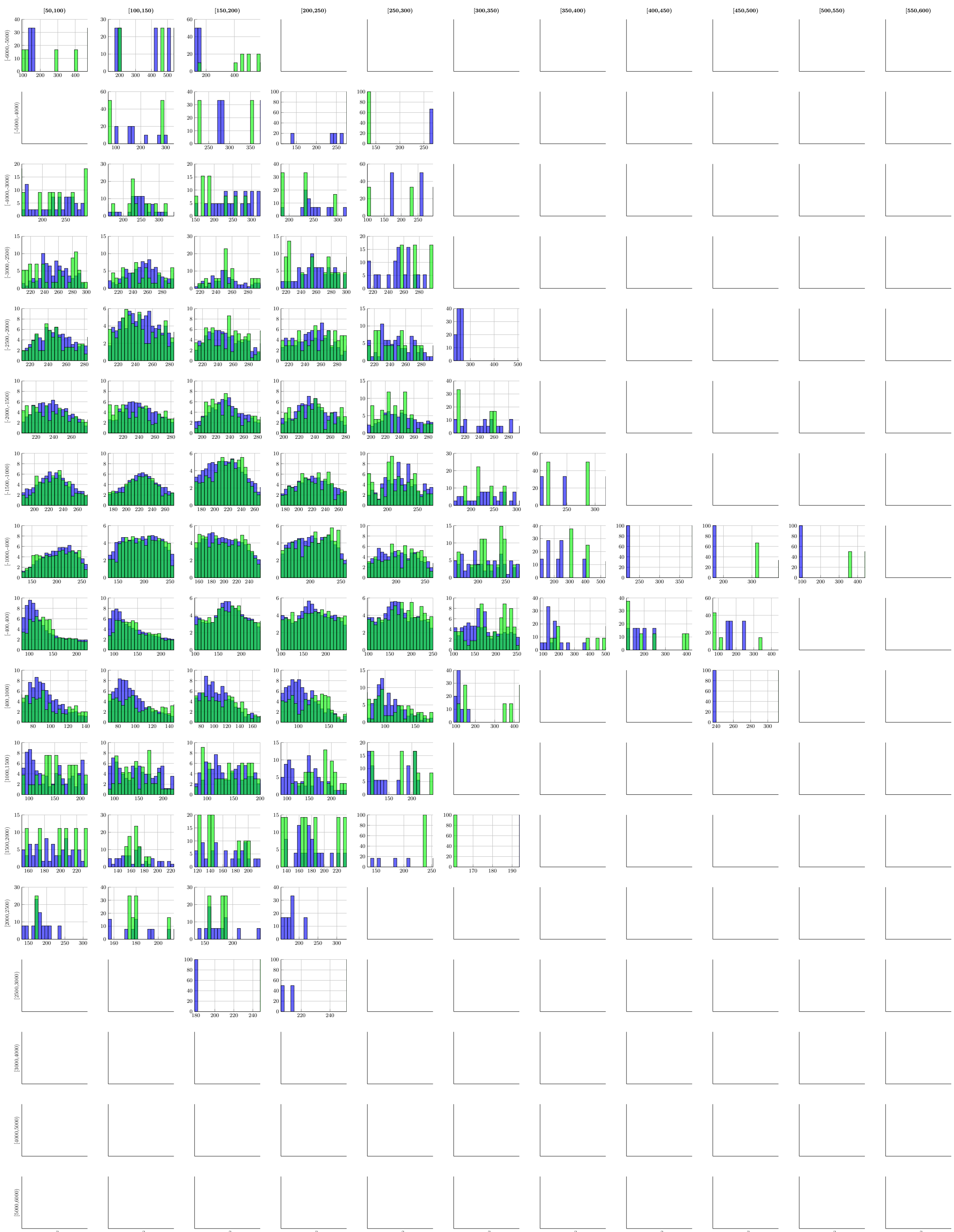

Figure F.13. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=2, C_1=Discrete code.

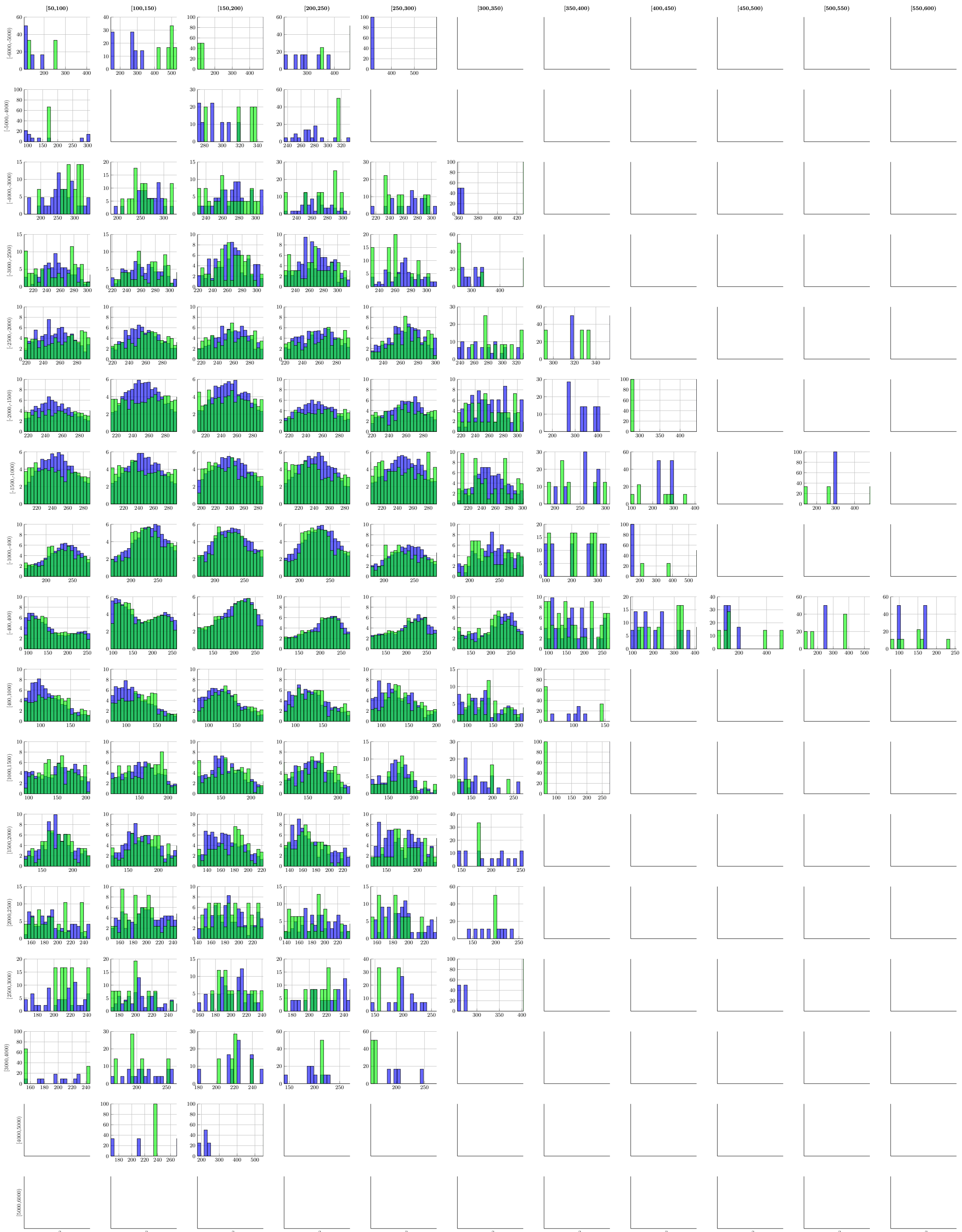

Figure F.14. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=3, C_1=Discrete code.

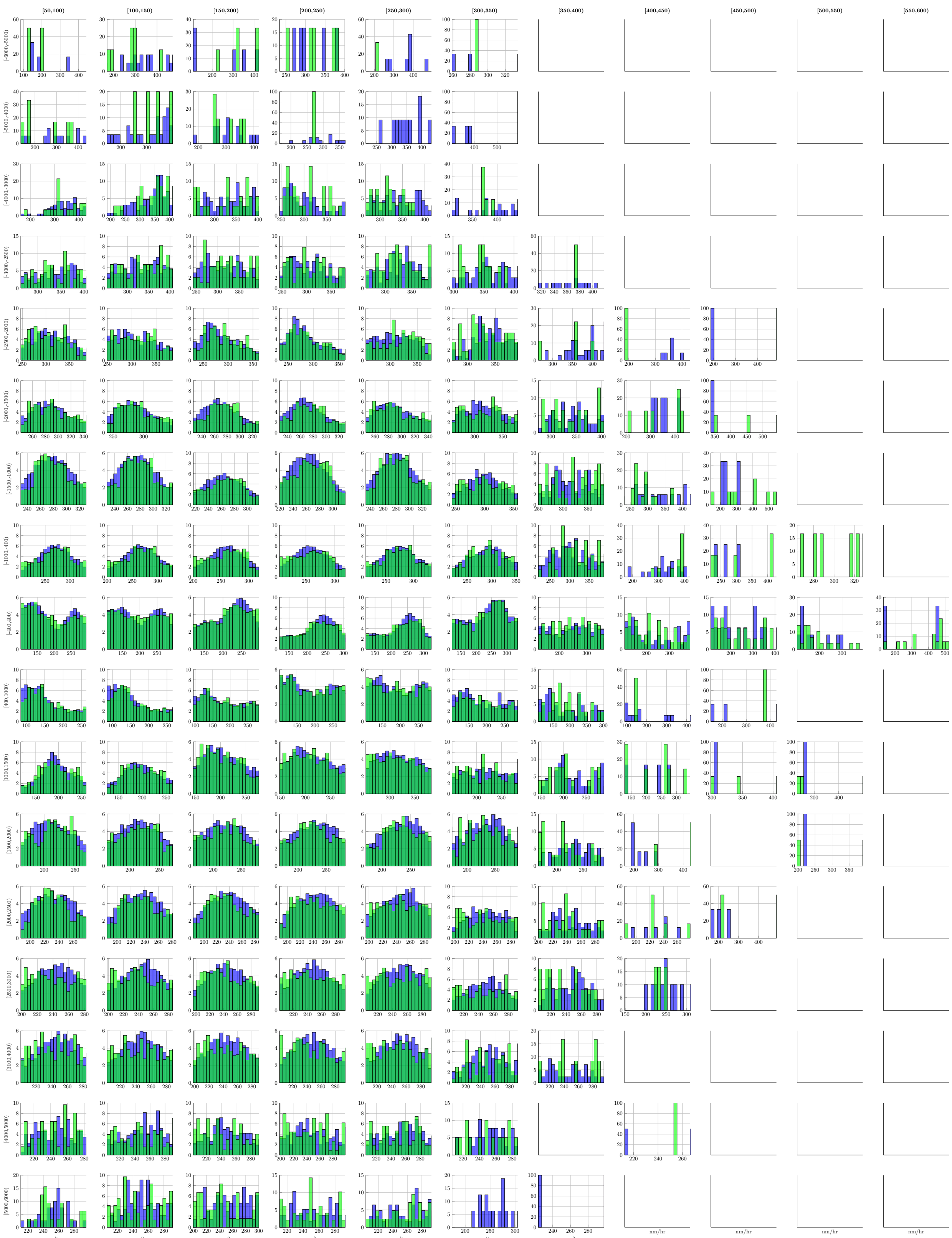

Figure F.15. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=4, C_1=Discrete code.

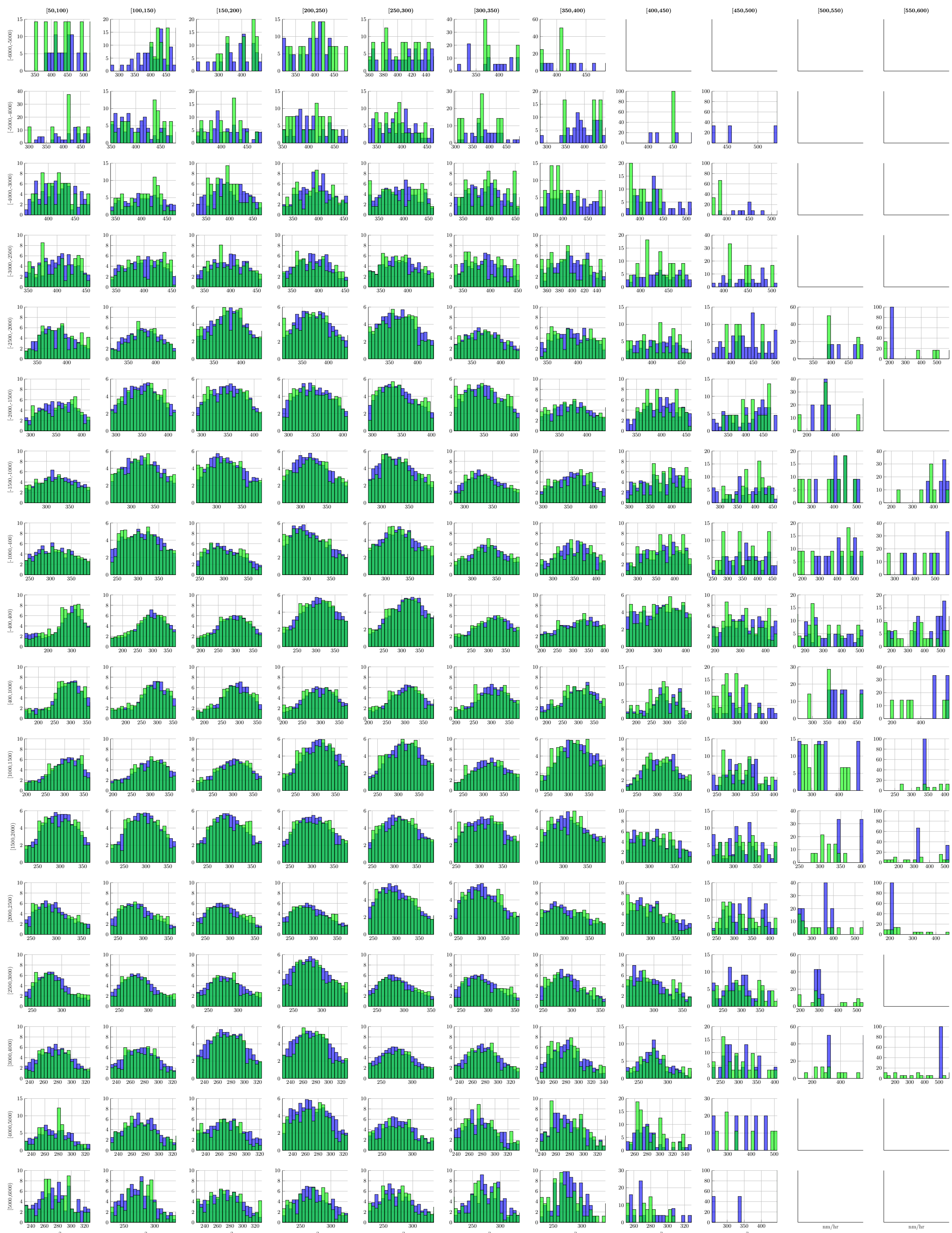

Figure F.16. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=5, C_1=Discrete code.

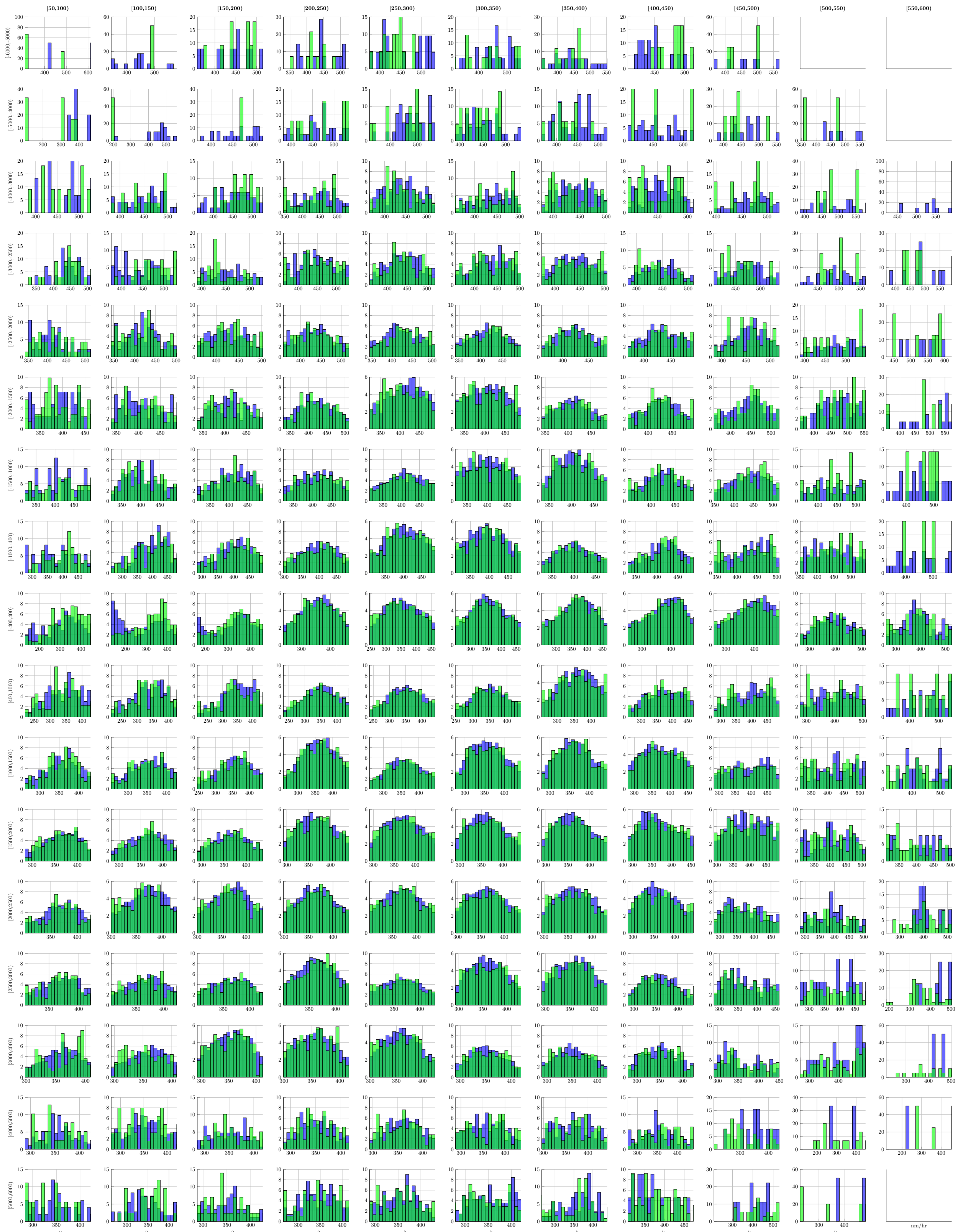

Figure F.17. The breakdown of v_1 by $\dot{h}_1$ and v_2 where L=6, C_1=Discrete code.

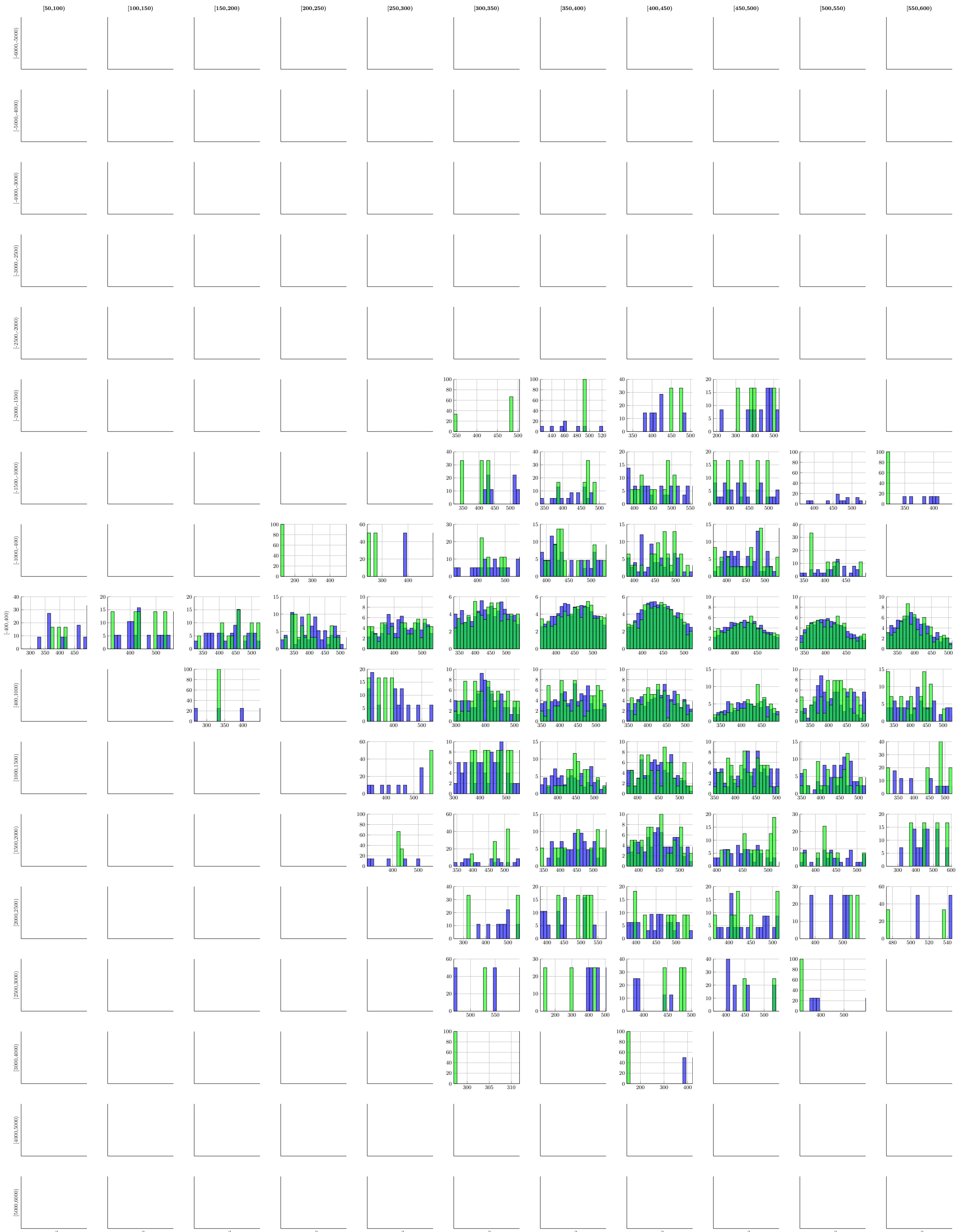

Figure F.18. The breakdown of υ_1 by $\dot{h}_1$ and υ_2 where L=8, C_1=Discrete code.

Figure F.19. The breakdown of $\dot{v}_2$ by $\dot{h_2}$ and v_2 where L=1.

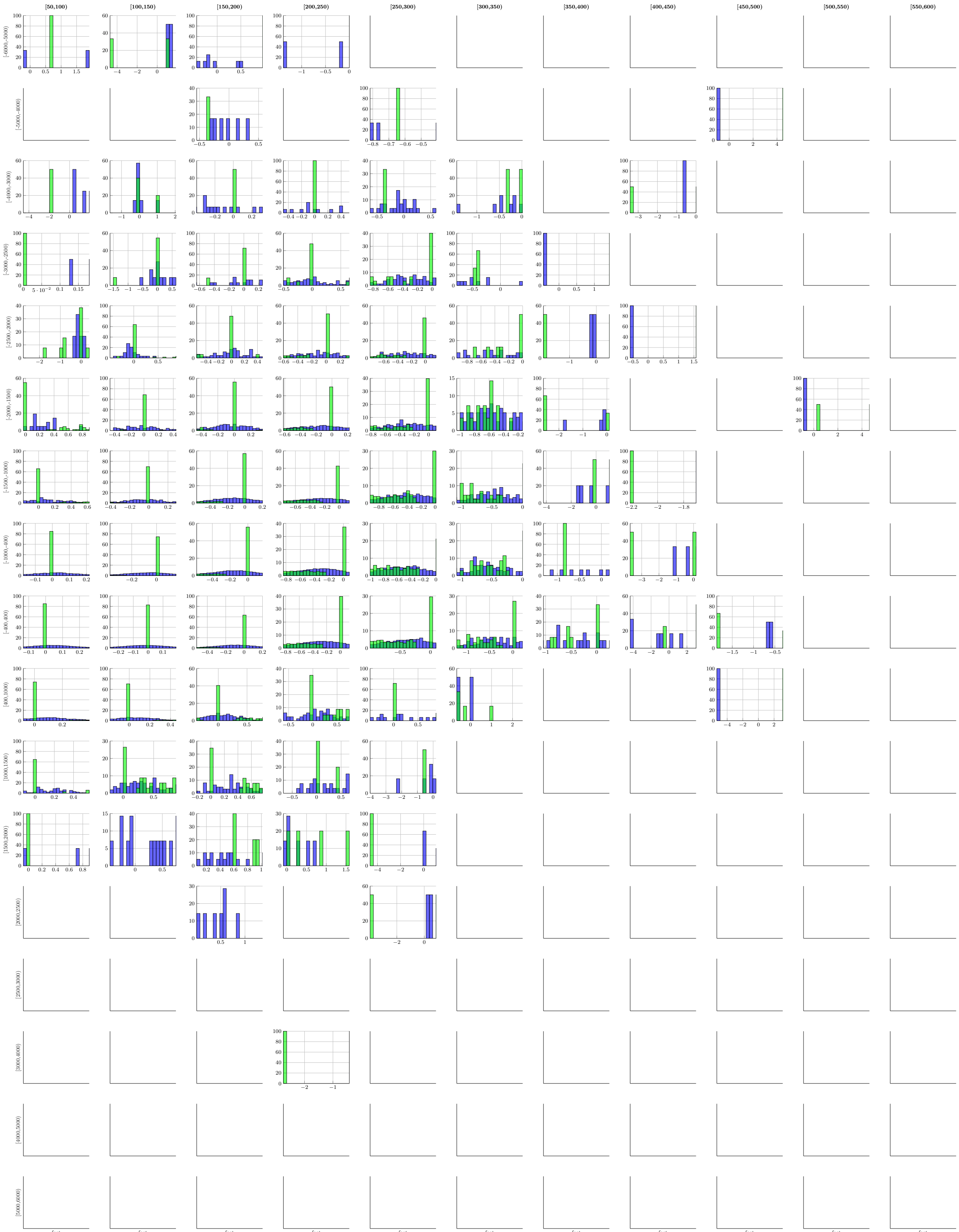

Figure F.20. The breakdown of $\dot{\upsilon}_2$ by $\dot{h}_2$ and υ_2 where L=2.

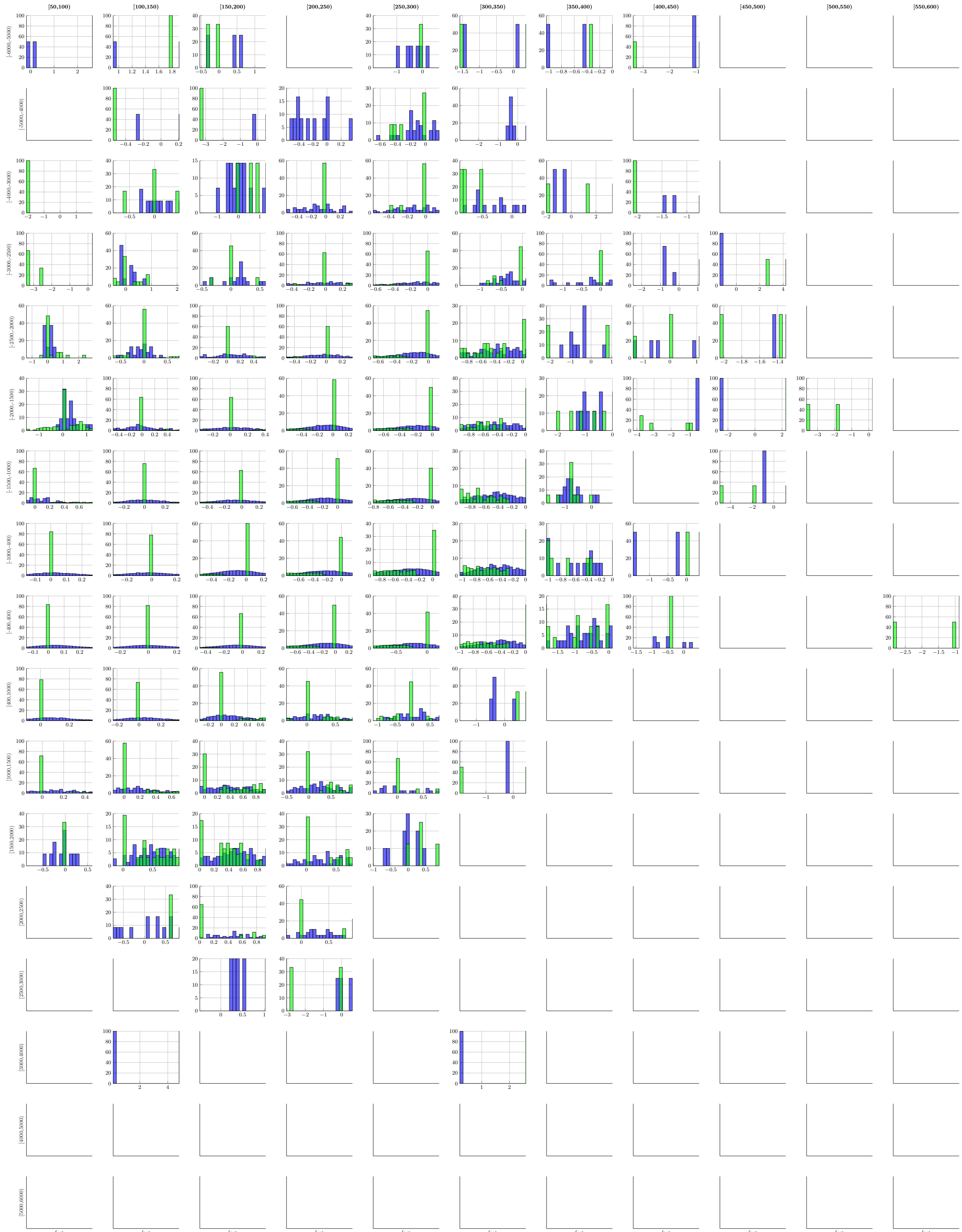

Figure F.21. The breakdown of $\dot{v}_2$ by $\dot{h}_2$ and v_2 where L=3.

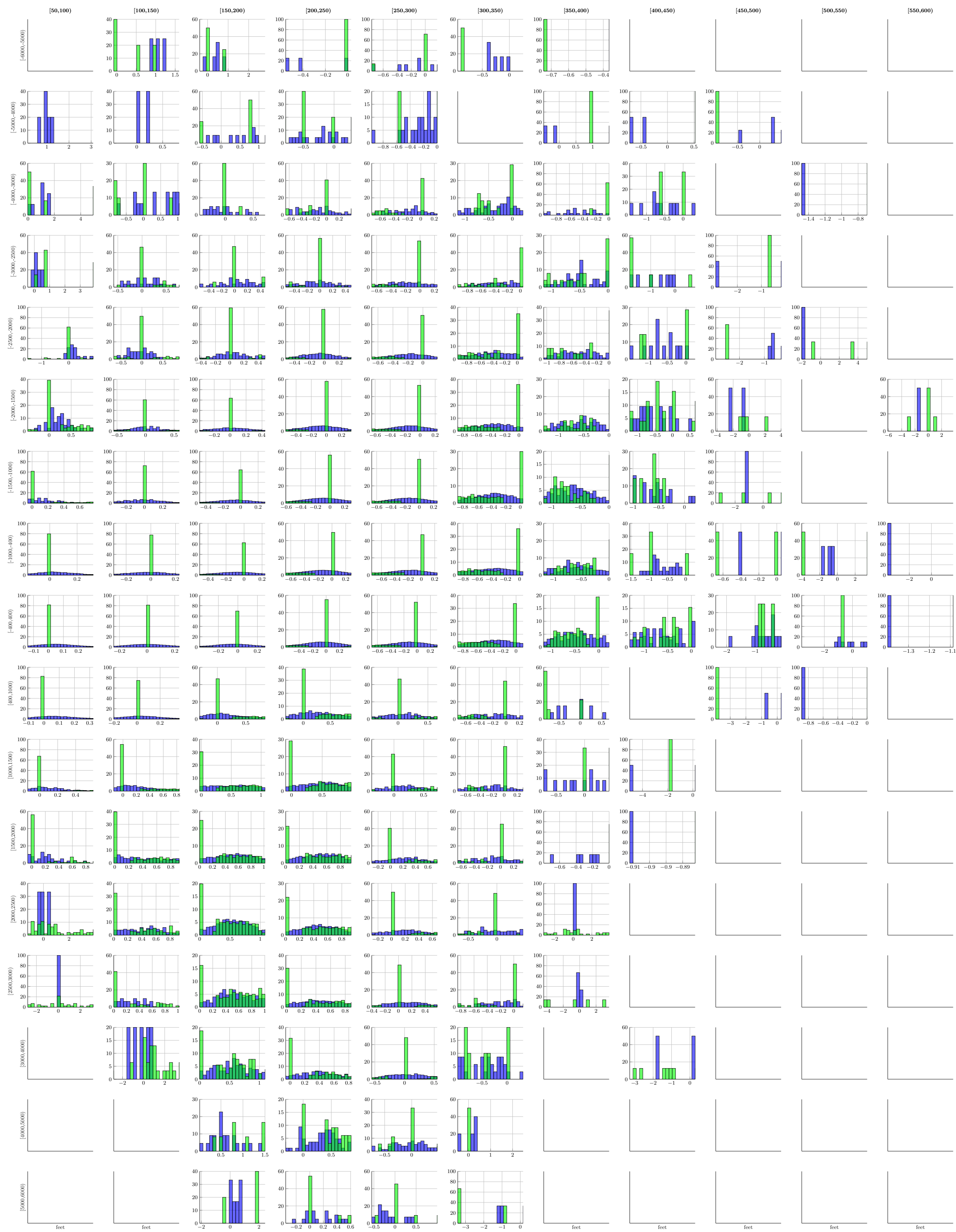

Figure F.22. The breakdown of $\dot{v}_2$ by $\dot{h}_2$ and v_2 where L=4.

This page intentionally left blank.

G. GLOSSARY

AC1 Aircraft 1

AC2 Aircraft 2

ACAS X Airborne Collision Avoidance System

ADS-B Automatic dependent surveillance-broadcast

AGL Above Ground Level

ARA Airborne Radar Approach

ARSR Air Route Surveillance Radar

ASR Airport Surveillance Radar

ATC Air Traffic Control

ATM Air traffic management

CEM 2008 Correlated Encounter Model for Cooperative Aircraft

CONUS Contiguous United States

CPA Closest Point of Approach

DAA Detect and Avoid

DHS Department of Homeland Security

DTED Digital Terrain Elevation Data

DOD Department of Defense

ECEM Extended Correlated Encounter Model

EUROCONTROL European Organization for the Safety of Air Navigation

FAA Federal Aviation Administration

FL Flight Level

HMD Horizontal Miss Distance

ICAO International Civil Aviation Organization

IFR Instrument Flight Rules

LLGrid Lincoln Laboratory Grid

MDP Markov decision process

MODSEF Mode S Experimental Facility

MOPS Minimum operational performance standards

MSL Mean Sea Level

MIT LL MIT Lincoln Laboratory

NAS U.S. National Airspace System

NASR National Airspace System Resources database

NGA Natioanl Geospatial Intelligence Agency

NHZ Non-Hazard Zone

NMAC Near mid-air collision

NM Nautical Miles

PCHIP Piecewise Cubic Hermite Interpolating Polynomial

RA resolution advisory

RADES Radar Evaluation Squadron

RTCA Radio Technical Commission for Aeronautics

SC-228 Special Committee 228: MOPS for UAS

SUA Special Use Airspace

SUAS Small unmanned aircraft system

TCA Time of Closest Approach

TCAS Traffic Alert and Collision Avoidance System

TFMS Traffic Flow Management System

USA United States of America

UAS Unmanned aircraft system

VFR Visual flight rules

REFERENCES

[1] SC-228, “Minimum operational performance standarsd (mops) for detect and avoid (daa) systems,” RTCA, Inc, Technical Report RTCA DO-365 (2017).

[2] M.P. McLaughlin, “Safety study of the Traffic Alert and Collision Avoidance System (TCAS II),” MITRE Corporation, Technical Rep. MTR 97W32 (1997).

[3] M. Corp, “System Safety Study of Minimum TCAS II,” MTR-83W241, McLean, VA., Technical rep. (1983).

[4] ICAO, “Acas manual,” Eurocontrol, Technical Report SCRSP/1-WP/53 (2004).

[5] K.C. T Arino, H.H. S Chabert, B.R. T Miquel, and E.V. K Rigotti, “Studies on the safety of acas ii in europe,” Eurocontrol, Technical Report ACASA/WP-1.8/210D (2002).

[6] M.J. Kochenderfer, J.K. Kuchar, L.P. Espindle, and J.D. Griffith, “Uncorrelated encounter model of the National Airspace System version 1.0,” MIT Lincoln Laboratory, Lexington, Massachusetts, Project Report ATC-345 (2008).

[7] A.J. Weinert, E.P. Harkleroad, J.D. Griffith, M.W. Edwards, and M.J. Kochenderfer, “Uncorrelated encounter model of the national AIrspace system version 2.0,” Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA, Project Report ATC-404 (2013).

[8] M.J. Kochenderfer, L.P. Espindle, J.K. Kuchar, and J.D. Griffith, “Correlated encounter model for cooperative aircraft in the national airspace system,” Massachusetts Institute of Technology, Lincoln Laboratory, Project Report ATC-344 (2008).

[9] M.W. Edwards, M.J. Kochendedrfer, J.K. Kuchar, and L.P. Espindle, “Encounter models for unconventional aircraft, version 1.0,” Massachusetts Institute of Technology, Lincoln Laboratory, Project Report ATC-348 (2009).

[10] J. Griffith, M.W. Edwards, R.M. Miraflor, and A. Weinert, “Due Regard Encounter Model Version 1.0,” Massachusetts Institute of Technology, Lincoln Laboratory, Project Report ATC-397 (2013).

[11] R.D. Grappel, “ASR-9 Processor Augmentation Card (9-PAC) phase II scan-scan correlator algorithms,” MIT Lincoln Laboratory, Project Report ATC-298 (2001).

[12] J.L. Gertz and R.D. Grappel, “Surveillance improvement algorithms for Airport Surface Detection Equipment Model X (ASDE-X) at Dallas-Fort Worth Airport,” MIT Lincoln Laboratory, Project Report ATC-333 (2007).

This page intentionally left blank.

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**			
1. REPORT DATE *(DD-MM-YYYY)* 08-05-2018	**2. REPORT TYPE** Project Report		**3. DATES COVERED** *(From - To)*
4. TITLE AND SUBTITLE Correlated Encounter Model for Cooperative Aircraft in the National Airspace System; Version 2.0			**5a. CONTRACT NUMBER** FA8721-05-C-0002 & FA8702-15-D-0001
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) N. Underhill, E. Harkleroad, R. Guendel, A. Weinert, D. Maki, and M. Edwards			**5d. PROJECT NUMBER** 1844
			5e. TASK NUMBER 11
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02421-6426			**8. PERFORMING ORGANIZATION REPORT NUMBER** ATC-440
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFLCMC/WING Building 557 1640 Loop Road West Wright Patterson AFB, OH 45433-7106			**10. SPONSOR/MONITOR'S ACRONYM(S)**
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release: distribution unlimited.			
13. SUPPLEMENTARY NOTES			

14. ABSTRACT

Detect and avoid (DAA) systems are designed to enable safe operation of unmanned aircraft systems (UAS) in the U.S. National Airspace System (NAS). Testing DAA systems before fielding is necessary to validate system robustness and confirm the system's safety benefit. The Extended Correlated Encounter Model (ECEM) generates encounters with trajectories beginning 110 seconds prior to Closest Point of Approach (CPA), enabling testing of the largest DAA alerting conditions that have been defined by the Radio Technical Commission for Aeronautics (RTCA). This model generates statistically representative encounters between two aircraft in the U.S. National Airspace System (NAS) in which both aircraft are cooperative (carrying a transponder) and at least one is receiving Air Traffic Control (ATC) services. Similar to the 2008 Correlated Encounter Model for Cooperative Aircraft (CEM), the ECEM uses a Bayesian network to capture encounter feature correlations and uses a Markov model to generate sensible and smooth aircraft maneuvers. The ECEM allows horizontal and vertical maneuvers and, because of the extended duration, it additionally allows acceleration variations. The ECEM also explicitly models airspeeds, vertical rates, turn rates, altitude layer, and airspace class enabling specific encounter subsets to be generated by limiting or specifying variable values in the encounter model structure.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	Same as report	138	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.